UNIVERSIDADE FEDERAL DO PARANÁ

MARIA JULIA CUNHA GARCIA

SCHOOLING EXPECTATION, RISK BEHAVIOR, AND BULLYING AMONG BRAZILIAN STUDENTS

CURITIBA

2021

MARIA JULIA CUNHA GARCIA

SCHOOLING EXPECTATION, RISK BEHAVIOR, AND BULLYING AMONG BRAZILIAN STUDENTS

Dissertação submetida como requerimento parcial para grau de Mestre em Desenvolvimento Econômico, no Programa de Pós-Graduação em Desenvolvimento Econômico, Setor de Ciências Sociais, Universidade Federal do Paraná.

Orientadora: Profª Drª Adriana Sbicca
Coorientadora: Profª Drª Kênia Barreiro de Souza

CURITIBA

2021

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS – SIBI/UFPR COM DADOS FORNECIDOS PELO(A) AUTOR(A)
Bibliotecário: Eduardo Silveira – CRB 9/1921

Garcia, Maria Julia Cunha
Schooling expectation, risk behavior, and bullying among brazilian students / Maria Julia Cunha Garcia.- 2021.
180 p.

Dissertação (Mestrado) - Universidade Federal do Paraná. Programa de Pós-Graduação em Desenvolvimento Econômico, do Setor de Ciências Sociais Aplicadas.

Orientadora: Adriana Sbicca Fernandes.

Coorientadora: Kênia Barreiro de Souza.

Defesa: Curitiba, 2021.

1. Desenvolvimento econômico. 2. Estudantes. 3. Comportamentos. 4. Riscos. 5. Bullying. I. Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas. Programa de Pós-Graduação em Desenvolvimento Econômico. II. Fernandes, Adriana Sbicca. III. Souza, Kênia Barreiro de. IV. Título.

CDD 302

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em DESENVOLVIMENTO ECONÔMICO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **MARIA JULIA CUNHA GARCIA** intitulada: **SCHOOLING EXPECTATION, RISK BEHAVIOR, AND BULLYING AMONG BRAZILIAN STUDENTS**, sob orientação da Profa. Dra. ADRIANA SBICCA FERNANDES, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 15 de Março de 2021.

Assinatura Eletrônica
16/03/2021 19:20:15.0
ADRIANA SBICCA FERNANDES
Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
16/03/2021 19:46:53.0
GILVAN RAMALHO GUEDES
Avaliador Externo (UNIVERSIDADE FEDERAL DE MINAS GERAIS)

Assinatura Eletrônica
16/03/2021 15:36:53.0
VICTOR RODRIGUES DE OLIVEIRA
Avaliador Externo (PROFESSOR PESQUISADOR- UFPR)

Este trabalho é dedicado a todos os jovens que sonham
e aspiram um futuro melhor.

AGRADECIMENTOS

Agradeço profundamente a minha família por todo o suporte e por sempre acreditarem em mim. Em especial, agradeço minha mãe Mirna da Cunha de Oliveira, o seu cuidado e amor são inigualáveis. Agradecimentos a todos os meus amigos que me incentivaram e contribuíram para esse trabalho de alguma forma. Agradeço também meus colegas de mestrado, a amizade deles foi o maior presente que a pós-graduação poderia me dar. Direciono agradecimento a Marco Rispoli, que foi meu suporte emocional durante maior parte dessa caminhada e me inspirou a querer ser uma pesquisadora.

Agradeço a minha orientadora Adriana e coorientadora Kênia por toda a paciência e interesse, seus ensinamentos serão levados para a vida e com certeza moldaram a minha identidade como pesquisadora. Agradecimentos especiais à banca, por aceitarem o convite e contribuirem para o estudo. Por fim, agradeço o Setor de Ciências Sociais Aplicadas e a Universidade Federal do Paraná pelo acolhimento durante esses dois anos e a Coordenação de Aperfeiçoamento de Pessoal em Nível Superior por proporcionar toda essa experiência via financiamento por bolsa de estudos.

"It is when we know our distributed choices are suboptimal that we become conscious of a process of decision making, one that is attuned to our personal irrationalities". (Herrnstein, 1992)

RESUMO

Comportamentos de risco podem surgir durante o período da adolescência, uma fase de descobertas e maturação da identidade, podendo haver consequências ao longo da vida do indivíduo. Nos três ensaios dessa dissertação se trabalha com jovens do nono ano do ensino fundamental, da base de dados da PeNSE de 2015, em que se pretende investigar aspectos como expectativa de escolaridade, comportamentos de risco e *bullying*. O primeiro ensaio investiga como a expectativa de escolaridade se relaciona com diversos aspectos socioeconômicos, relacionais e de risco. Estatísticas descritivas e modelo logístico ordenado demonstraram que jovens em condição socioeconômica vulnerável, com problemas relacionais e associados a comportamentos de risco têm uma maior prevalência de expectativas de escolaridade menores. Tais resultados podem indicar possíveis caminhos adequando mensagens e ações a públicos mais sujeitos à menor expectativa. No segundo ensaio, são analisadas as relações entre o envolvimento dos jovens em atividades de risco e as expectativas de escolaridade. Considera-se que inconsistências e vieses comportamentais podem servir de gatilho para a prática de atividades de risco que possuem influências negativas no bem-estar. A metodologia usada é a de modelos de equações estruturais, que permite a estruturação de construtos para comportamento de risco alcoólico, sexual e uso de substâncias, mensurando o nível de engajamento em tais atividades. Os resultados demonstraram a existência de associações entre comportamentos de risco e expectativa de escolaridade, embora o direcionamento de tais relações sejam distintas a depender do tipo de risco abordado. Comportamentos como o uso de substâncias, tendem a não ter efeito sobre as expectativas, indicando a necessidade de políticas que promovam uma maior conscientização sobre os custos relacionados ao engajamento em tais atividades. Além disso, diversos mediadores são testados para verificação do efeito de variáveis externas ao modelo de mensuração. O terceiro ensaio trata do *bullying*, um tipo de violência que pode ter implicações perversas em diversos aspectos da vida dos jovens que o praticam e que são vítimas, sendo uma delas o engajamento em comportamentos de risco. Por meio de uma análise afetiva de risco, considerando a heurística do afeto, este estudo buscou verificar como a vitimização por *bullying* pode impactar os comportamentos alcoólico, sexual, de uso de substâncias e a relação com pares de risco. Utiliza-se análise fatorial exploratória e método de pareamento por escores de propensão para identificar o efeito da prática e da vitimização por *bullying* nos comportamentos de risco, também considerando a implicação por diferentes tipos de vitimização. Os resultados demonstraram que há uma prevalência no impacto positivo do *bullying* nos comportamentos de risco, com maiores magnitudes para jovens que praticam e que são vitimizados em razão da sua orientação sexual. Por fim, enfatiza-se que redes de proteção e compreensão da análise afetiva de risco são fatores importantes para estabelecimento de políticas que visem reduzir o *bullying* e, assim, os comportamentos de risco, sendo necessária a integração de profissionais da saúde no contexto social escolar.

Palavras-chaves: Comportamento de Risco. Expectativa de Escolaridade. Bullying.

ABSTRACT

Risk behaviors may arise during adolescence, a phase of discovery and identity maturation, which may have consequences throughout life. We developed three essays focusing on students from the ninth year of elementary school, from the 2015 PeNSE database. We investigate aspects such as schooling expectations, risk behavior, and bullying. The first essay investigates how schooling expectations are related to various socio-economic, relational, and risk aspects. Descriptive statistics and ordinal logistic models demonstrate that students in vulnerable socio-economic conditions, with relational problems, and associated with risky behaviors have a higher prevalence of low-schooling expectations. The second essay analyzes students' involvement in risk activities and schooling expectations relations. It is considered that inconsistencies and behavioral biases may serve as a trigger for the practice of risky activities that have negative influences on well-being. We used structural equation models, which allow the construction of latent variables for alcoholic, sexual, and substance use risk behavior, measuring the engagement in such activities. The results demonstrated associations between risk behaviors and educational expectations, although such relationships' directions are different depending on the type of risk. Behaviors such as substance use did not affect expectations, indicating the need for policies that promote greater awareness of the costs related to risk engagement. Also, several mediators are tested to verify the effect of variables external to the measurement model. The third essay deals with bullying, a type of violence that can have perverse implications on students involved practicing or as a victim. Affective risk analysis considers the affect's heuristic. This study sought to verify how victimization by bullying may impact alcoholic, sexual, substance behaviors, and peer relations. Exploratory factor analysis and propensity score matching are applied to identify the effect of practice and victimization by bullying on risk behaviors. Also, different types of victimization are tested. The results showed a prevalence of the positive impact of bullying on risk behaviors, with greater magnitudes for students who practice and are victims due to their sexual orientation. Finally, it is emphasized that protection networks and an understanding of affective risk analysis are relevant to establish policies aiming to reduce bullying and, thus, risk behaviors. Such an objective requires the integration of health professionals in the school social context.

Key-words: Risk behavior. Schooling Expectation. Bullying.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

GENERAL INTRODUCTION

Adolescence represents the transition to adulthood, and it may be a phase of discoveries and changes. At this moment, youth may not have enough knowledge to decide under relevant circumstances or be easily convinced by peers and family pressure. Despite it, they still need to decide on important subjects that could potentially impact their lives. A relevant decision regarding a regular adolescent is related to his educational prospect. This aspect may generate significant effects once more educated people are likely to have better jobs and higher earnings. Still, this decision may be influenced by other aspects, with risk-taking being one of them.

Young people might engage in risky activities for several reasons, such as seek for novelty, intensity, peer influence, and the need for acceptance. Further, some features that arise in a school ambiance may be decisive to this engagement. This dissertation aims to analyze the aspects of student's life related to their decisions on schooling expectations and risky behaviors. We used subjective data to investigate the role of schooling expectation and how this prospect can be associated with other socioeconomic, relational, behavioral, and risk variables. Also, the analysis focus on the relationship between risky behaviors and schooling expectations. Finally, we investigate bullying victimization related to risk behaviors for victims and aggressors.

Sustained by a set of theoretical backgrounds, with attention to behavioral economics, this dissertation addresses important features on students' transition to adulthood and its impact on their lives in many ways. Our focus is to analyze the results taking into account empirical evidence and practical policies applicable to the students' context.

To reach this objective, the database used is the third edition of the National School-Based Health Survey [1] (PeNSE) held in 2015. Using a digital device, the students answered 127 questions about a large set of information, such as school, social, health, risk, and family aspects. PeNSE survey was developed by the Brazilian Institute of Geography and Statistics (IBGE[2]) aligned with the Ministry of Health and supported by the Ministry of Education. Besides, the sample plan comes from the 2013 School Census.

Data was divided into two sample plans. The first sample plan covers students attending the 9^{th} grade of elementary school-aged 13 to 15. It was implemented in 3,160 schools, 4,159 classes, with 102,301 students participating, accounting for 85.2% of 120,122 individuals initially eligible for the study. The second sample focused on students, from 13 to 17 years old attending 6^{th} grade of elementary school till 3^{rd} grade of high school. This one covered 380 schools with 653 classes, reaching 16,556

[1] The comments addressed in this section about PeNSE can be found in the methodological notes of the survey, to see more (IBGE, 2016)

[2] From Portuguese: Instituto Brasileiro de Geografia e Estatística

students from 16,608 originally selected. These stages were chosen because, at this age, these students already have enough knowledge to understand and answer a self-administered questionnaire.

Moreover, regarding regional coverage, the first sample is wider, representing capitals, metropolitan regions, states, a federal district, and Brazil. At the same time, the second sample represents metropolitan regions and the country. Therefore, based on coverage and sample size, this dissertation uses the first sample to analyze the subjects of schooling expectation, risky behavior, and bullying among students. Additionally, it is worth mentioning that PeNSE 2015 was approved by the National Research Ethics Commission (Conep) of the National Health Council (CNS), which regulates and approves health research involving human beings, through Conep Report n. 1,006,467, of March 30, 2015.

Some limitations in this study must be discussed. First, the age groups' representativeness considers mostly students aged 13 to 15 years old, with older groups being found as a residual. Second, this questionnaire has not been subjected to psychometric validation in Brazil, as far as we know. Third, some inconsistencies[3] in the questionnaire can be found, especially concerning alcohol risk behavior. Fourth, the survey target is students, so the sample is likely to be biased towards risk-averse people.

The inconsistencies presented in the survey were discarded, and we relied mostly on the test results to confirm the validation, but little can be done regarding the first and fourth limitations. Therefore, we restrict the validation of our results to the youth population attending a school that still represents a large set of Brazilian young people. Lastly, PeNSE has the largest dataset about young students in Brazil and has been the source of several studies concerning this population, including practical implications to address policies to solve vulnerabilities. We hope this dissertation may contribute to the literature focusing on the role of schooling expectation, risky behaviors, and bullying victimization.

This dissertation is structured as follows. The first chapter analyzes schooling expectations and associated factors. The second chapter investigates the role of risky behaviors on schooling expectations. In the third chapter affective risk analysis is used to verify the impact of bullying on risk behaviors. At last, the conclusion draws the main analysis of this dissertation.

[3] The inconsistencies will be discussed in the data subsection - Chapter one.

1 AN EXPLORATORY ANALYSIS OF SUBJECTIVE SCHOOLING EXPECTATIONS OF BRAZILIAN STUDENTS

1.1 INTRODUCTION

Human capital investment generally acts as a promotive factor for productivity improvement and to raise earnings. In such a case, the investment in education is understandably seen as one of the major factors in turning around deep and persistent disadvantages (COMBER, 2014). Unfortunately, in many developing countries with relatively high-income levels, the stock of education has grown slowly, despite high returns to acquiring education (ATTANASIO; KAUFMANN, 2014).

The empirical literature has demonstrated that some characteristics, such as parents' education, socioeconomic status, and attendance, may be behind much of the schooling-growth connection (GLAESER, 1994). Given the importance of the returns to schooling in economics efforts to study human capital investment, it might be critical to make substantial efforts to comprehend the youth's expectations. Moreover, the schooling expectation is significant to schooling decisions, once people evaluate their options and possible outcomes (GIUSTINELLI; MANSKI, 2018).

The use of subjective data is not commonly seen in economics but has been expanding in recent decades. It is worth mentioning that considering an expectation does not proxy later attainment, but it can indicate the motivation for distinctly high educational attainment (LAUGLO; LIU, 2019). Besides, data on subjective expectations can address a relevant question regarding the need for *ad hoc* assumptions about information and formation of expectations (ATTANASIO; KAUFMANN, 2014).

Based on the relevance that subjective data on schooling expectations have, this paper investigates the role of schooling expectations and how they are related to socioeconomic, behavioral, and risk variables. Descriptive statistics and ordinal logistic regressions are applied to verify the associations related to a higher schooling expectation considering the 2015 National School-Based Health Survey (PeNSE). In this data, students were asked, "What is the highest level of education that you intend to conclude?". Based on this question, the analysis is conducted to understand the main factors related to schooling expectations among Brazilian students.

This paper has two contributions. First, we analyze the role of expectations and how they can be related to other dimensions of students' lives. Second, we used a methodology that, to the best of our knowledge, has not been applied to the PeNSE database. The article is structured in four sections, beyond this introduction. The first section reviews the literature. The second presents the methodology applied. The third section addresses the results. Fourth, a discussion is carried out, aggregating the main discoveries in the literature. The fifth section concludes.

1.2 LITERATURE REVIEW

Investments in human capital involve expenditures on education, training, and other dimensions that might work to provide youths' knowledge, skills, and a way of analyzing problems (BECKER, 1993). As the schooling decisions have significant implications to the future of youth, the willingness to analyze subjective data on economic analysis of schooling expectations constitutes an important step to understand the perspective that students hold regarding their future (MANSKI, 1993; GIUSTINELLI; MANSKI, 2018).

The standard practice in economic analysis is to use revealed preferences, with expectations being implicit as assumptions. However, to understand choice analysis, the integration of interpretable subjective data aligned with choice data represents a strong standpoint to a more structured analysis (MANSKI, 1993; MANSKI, 2004). Wilson, Wolfe e Haveman (2005) developed a model with rational utility maximizers assuming that youth respond to expectation based on the available options. In this model, a young person values consumption and schooling with various background, family, and community variables attached to schooling and consumption weights. This function is represented by:

$$U_{si} = B_s X_i + Z C_{si} + \epsilon_{si} \tag{1.1}$$

In equation 1.1, the utility of a youth *i* with its school level s is given by X_i, which is a vector of nonconsumption variables that affect utility from schooling, such as background, family, and community. B_s represents weights attached to nonconsumption utility gains of schooling. C_{si} describes a lifetime discounted stream of consumption conditional on schooling expectation. Z is the weight of consumption in utility, and $\epsilon_s i$ is a random utility conditional on schooling level.

Allowing the utility effect of schooling to differ among youths based on their socio-economic background, Wilson, Wolfe e Haveman (2005) assume those characteristics may affect youths' perceptions. Moreover, the model presumes that youth do not know their future income prospect with certainty and form expectations based on the income of older youths with similar characteristics. Such aspect is aligned to what Manski (1993) proposed on how young people process information. To his understanding, if youth use data on realized outcomes to create their expectations, then the data depend on how they think other youths make schooling decisions. The only possibility to a homogeneous expectation among youth would be if they observe the same behavior in all of their peers. The Wilson, Wolfe e Haveman (2005) assumption that they based their expectation on people with similar characteristics implicitly assumes that a wide range of heterogeneity on schooling expectations is due to the heterogeneity of peers.

This analysis depends on what reference groups are underlined, the youth will choose to graduate if the utility from graduating at a school level is greater than the expected utility from not graduating. This paper aims to analyze the X_i's features of

equation 1.1, to verify which variables could be more related to a higher schooling expectation and, consequently, which variables could be more related to a lower utility on schooling.

Furthermore, this analysis demonstrates how social capital might play an important role in human capital investment. As stated by Coleman (1988), this fashion of capital, more related to the communicational and relational dimension, can reinforce and generate better schooling results, being as important as human and physical capital with relevant ramifications on future generations. Additionally, an important consideration from this analysis is that youth learn through observation, even though, as stated by Manski (1993), they cannot observe outcomes that people with similar characteristics would have experienced once they had made other decisions. So, the outcomes from counterfactual scenarios also remain a source of prospection in the schooling expectations of youths.

The literature related to schooling expectations and achievements draws attention to a varied set of drastic transitions even in the first years of schooling (SEEFELDT; GALPER; DENTON, 1997). Wilson, Wolfe e Haveman (2005), with a cohort study that followed 1,942 young people for 20 years, found that income returns expectation from high school was influential in the youth schooling choices. The result remained even after controlling for background variables. Heard (2007), based on ordinary least squares and logistic regression for 11,318 adolescents, verified that teenagers with non-two biological parents were less likely to have high schooling expectations.

Some differences can be addressed when considered findings by sex. Lauglo e Liu (2019), based on data of 50 countries, demonstrated a pro-girl disparity in adolescents' subjective expectation of higher education. However, this disparity may remain only in the expectation, and not attainment. (MELLO, 2008), using hierarchical models on a sample of 10,364 adolescents, identified that males who reported higher occupational expectations had higher attainment in adulthood. At the same time, female's occupational attainment did not vary by their prior expectations, though males and females had the same expectation average to go to college. Furthermore, Attanasio e Kaufmann (2014) studying 23,000 students, addressed that schooling decisions are particularly important for boys, based on the result that market expectation and monetary returns were strong predictors of their future attainment, with lower magnitudes for females.

A relevant aspect concerning expectation is raised by Smyth (2020), he identified that a reasonable proportion of the youth holds lower expectations than their parents and that higher education expectations were lower than the actual progression rate found among secondary students. This means that many students who aspire to high educational levels might get frustrated by not reaching them. Moreover, a lower prevalence of risk behaviors, such as smoking and drinking, can be related to higher schooling expectations. As Sutherland e Shepherd (2001) identified, from the students who planned to go to college, only 55.6% reported using any substances compared

with 73.5% among those who did not plan to continue studying after school, the study had 4,516 adolescents.

These studies generally support the importance of human capital investments, as well as social capital. A consequence of these decisions might be dropping out of school, and these dynamics are usually related to socioeconomic aspects (WEINSTEIN; MADISON; KUKLINSKI, 1995). This study aims to analyze how socioeconomic, relational/behavioral, and risk features may be related to a higher prevalence of schooling expectations. Identifying factors deemed key predictors of a lower schooling expectation may address paths to analyze aspects that can negatively affect youths' lives.

1.3 DATA AND METHODOLOGY

PeNSE data has a large number of information concerning students' health and behavior. Based on the literature, the variables were chosen to analyze the schooling prospect and verify how these variables can be associated with the student's expectations. In this section, analysis firstly demonstrates how the data was treated. After that, the model is presented.

1.3.1 Data

The PeNSE data is designed as a complex survey, allowing weights and strata in the research design. A survey package from the statistical software R was used (LUMLEY, 2004; LUMLEY, 2010). The data expansion was applied for all primary and secondary unit samples and geographical strata. Before that, missing data were discarded, and the sample decreased from 102,072 valid observations to 95,047. Despite it, sample proportions were not highly altered, except for alcohol[1] variables that have altered 1.7 percentage points at maximum, as shown in Tables 49 to 55 from Appendix A. All other variables did not reach one percentage point variation in the sample proportions.

With sample expansion, total observations represent 2,443,517 observations. The composition demonstrates that female students are in a higher proportion (52.01%), as well as those living in urban areas (77.02%), studying in public schools (85.16%), from the southeast region (43.46%), and living with their mother (90.05). However, the proportion of those who live with the father in the house was much smaller (63.62%). Some other socioeconomic results can be seen in Table 1. It is worth mentioning that, as the sample is focused on ninth-grade students, older students are represented at the margin, with 88.98% of them between 13 to 15 years old.

[1] This set of variables has a research inconsistency allowing students that respond "never drink alcohol" to respond to other alcohol variables, resulting in conflicting answers that were discarded.

TABLE 1 – Socioeconomic weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100.00	
Gender			
Girls	1,270,914	52.01	19,218
Boys	1,172,603	47.99	17,871
Age			
11	552	0.02	286
12	8,879	0.36	832
13	439,070	17.97	15,278
14	1,255,445	51.38	20,208
15	479,821	19.64	10,945
16	172,302	7.05	4,981
17	59,729	2.44	2,772
18	16,883	0.69	1,256
19	10,837	0.44	922
City			
Interior	561,450	22.98	9,577
Capital	1,882,067	77.02	32,155
Region			
North	235,016	9.62	5,700
Northeast	676,924	27.70	13,940
Southeast	1,062,064	43.46	28,475
South	287,730	11.78	8,098
Midwest	181,783	7.44	4,739
School type			
Private	362,624	14.84	24,606
Public	2,080,893	85.16	46,022
Skin Color			
White	886,044	36.26	17,444
Black	320,550	13.12	8,014
Yellow	100,535	4.11	3,320
Brown/mixed-Skin	1,056,273	43.23	19,447
Indígenous	80,115	3.28	3,016
Has a Cellphone			
Yes	2,141,846	87.65	30,352
No	301,672	12.35	7,435
Has Internet			
Yes	1,906,024	78.00	29,735
No	537,494	22.00	11,774
Has a Car			
Yes	1,385,344	56.69	22,597
No	1,058,173	43.31	20,078
Housekeeper			
Yes	230,056	9.41	9,016
No	2,213,462	90.59	33,579
Lives with the Mother			
Yes	2,200,273	90.05	30,878
No	243,244	9.95	5,377

Table 1 continued from previous page

Variable	Frequency	%	Error
Lives with the Father			
Yes	1,554,881	63.63	22,775
No	888,637	36.37	15,357
Number of People in the Residence			
1	4,081	0.17	652
2	136,292	5.58	3,890
3	470,095	19.24	8,416
4	817,496	33.46	13,129
5	527,305	21.58	10,181
6	254,575	10.42	6,963
7	110,900	4.54	3,775
8 or more	122,774	5.02	4,044

Source: Elaborated by the author using PeNSE data.

Relational and behavioral variables are considered in Table 2. They demonstrate that 26.11% of students had at least one smoking parent, that 23.18% had missed class without parents' permission in the recent 30 days, and students usually consider themselves as normal regarding their weight. Moreover, the human papillomavirus vaccination (HPV) was taken by 38.72% of the students. It is worth noting that the public health system offers this vaccine for free in this same age range, so this proportion is very low.

TABLE 2 – Relational/behavioral weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100.00	
At least one of the parents smoke			
Yes	637,945	26.11	12,962
No	1,805,572	73.89	25,061
Missed class without parental permission			
Yes	566,372	23.18	13,106
No	1,877,145	76.82	25,276
Body Perception			
Very Thin	128,716	5.27	3,864
Thin	498,726	20.41	10,091
Normal	1,365,473	55.88	18,803
Fat	398,488	16.31	7,686
Very Fat	52,114	2.13	2,153
Feels Lonely			
Yes	405,385	16.59	8,392
No	2,038,133	83.41	28,324
HPV Vaccination			
Yes	946,131	38.72	16,361
No	1,497,387	61.28	21,394
Parents Understand their Problems			
Yes	1,072,155	43.88	16,534

Table 2 continued from previous page

Variable	Frequency	%	Error
No	1,371,363	56.12	21,062
Well-Treated by the Colleagues			
Yes	1,529,281	62.59	21,568
No	914,236	37.41	16,391
Practiced Vitimization			
Yes	479,884	19.64	10,172
No	1,963,633	80.36	27,406
Suffered Victimization			
Yes	577,750	23.64	10,399
No	1,865,767	76.36	26,436
Suffered Bullying			
Yes	1,165,393	47.69	18,889
No	1,278,124	52.31	18,536
Number of Close Friends			
None	101,879	4.17	3,554
1	152,064	6.22	4,291
2	308,998	12.65	7,849
3 or more	1,880,576	76.96	25,047
Friends who drink alcohol			
None or don't know	681,122	27.87	13,862
Few	682,439	27.93	10,950
Some	576,961	23.61	9,893
Most of	432,936	17.72	9,007
All	70,060	2.87	3,207
Friends who use Drugs			
None or don't know	1,531,781	62.69	22,055
Few	482,086	19.73	10,501
Some	290,435	11.89	7,714
Most of	120,068	4.91	5,167
All	19,148	0.78	1,665
Victim of sexual violence			
Yes	93,453	3.82	3,465
No	2,350,064	96.18	32,456

Source: Elaborated by the author using PeNSE data.

Regarding the relational and affective variables, 16.59% have felt alone in the recent 12 months, a large proportion of them (56.12%) considered that their parents did not understand their problems, and was involved in bullying dynamics sometimes in life (47.69%) or the recent 30 days (23.64%). Also, they usually have 3 or more close friends (76.96%), friends who drink alcohol (72.13%), or use drugs (37.31%)[2] . Furthermore, being a victim of sexual violence is reported by 3.8% of the sample, which indicates that youth's safety is an important subject to Brazilian policy.

Analyzing PeNSE risk data, alcohol variables (Table 3) indicate that more than half of students (53.88%) have tried alcohol at least once, but frequency and intensity in

[2] This number represents the sum from few to all of Table 2

the recent 30 days were quite smaller. They have been drunk at least once in their lives for 21.46% of the sample, 6.97% alcohol drinking problems, and the main fashion to get the alcohol was in parties (13.42%) and with friends (5.50%).

TABLE 3 – Alcohol risk weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100.00	
Drank alcohol at least once			
Yes	1,316,689	53.88	21,860
No	1,126,828	46.12	17,681
Age started to drink alcohol			
Never drank	1,126,828	46.12	17,681
7 to 8	66,827	2.73	2,598
9 to 10	110,617	4.53	4,338
11 to 12	314,319	12.86	7,509
13 to 14	677,783	27.74	12,677
15 to 16	134,543	5.51	3,668
17 to 18	12,601	0.52	971
Frequency drank alcohol in the recent 30 days			
Never drank	1,126,828	46.12	17,681
None	731,233	29.93	13,181
1 or 2 days	353,939	14.48	7,609
3 to 5 days	100,860	4.13	3,383
6 to 9 days	57,447	2.35	2,375
10 to 19 days	42,823	1.75	2,203
20 or more days	30,386	1.24	1,737
Intensity of alcohol use in the recent 30 days			
Never drank	1,126,828	46.12	17,681
Did not drink	669,243	27.39	12,403
1 Glass or less	337,174	13.80	7,777
2 to 3 glasses	151,328	6.19	4,076
4 or more glasses	158,945	6.50	4,874
Number of times got really Drunk			
Never drank	1,126,828	46.12	17,681
None	792,250	32.42	13,860
1 to 2	340,889	13.95	7,571
3 to 5	96,025	3.93	3,584
6 to 9	28,265	1.16	2,121
10 or more	59,260	2.43	2,954
How did you acquired acohol in the recent 30 days			
Never drank	1,126,828	46.12	17,681
Did not drink	571,633	23.39	10,929
Store or Bar	106,170	4.34	3,740
Hawker	12,160	0.50	1,053
Someone bought it for me	26,764	1.10	1,777
Friends	134,415	5.50	3,825
Got it hidden from someone	28,075	1.15	1,729
From a person older than me	69,488	2.84	2,797
Party	327,948	13.42	7,988

Table 3 continued from previous page

Variable	Frequency	%	Error
Got it from another way	40,037	1.64	1,923
Times that had problems due to alcohol drinking			
Never drank	1,126,828	46.12	17,681
None	1,146,315	46.91	18,950
1 or 2 ever	105,960	4.34	3,711
3 to 5 ever	29,872	1.22	1,698
6 to 9 ever	10,002	0.41	954
10 or more ever	24,540	1.00	1,490

Source: Elaborated by the author using PeNSE data.

Cigarette use by students (Table 4) was verified in 18.32% of the sample, with frequency in the recent 30 days being reported by 5.53% of the sample. The cigarette acquisition dynamics is different from alcohol. From those who smoked in the recent 30 days, most of them got the cigarettes in stores/bars (1.94%), asked someone (1.51%), or got it hidden from someone (1.31%). Although cigarette selling for under 18 is forbidden under Brazilian law, of the 4.71% of the sample that has tried to buy cigarettes, 3.14% was able to buy cigarettes with no problems.

TABLE 4 – Cigar risk weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100	
Smoked Cigarette Once			
Yes	447,733	18.32	10,170
No	1,995,785	81.68	27,741
Age started to smoke			
Never Smoked	1,995,785	81.68	27,741
7 to 8	36,871	1.51	1,897
9 to 10	41,975	1.72	1,973
11 to 12	107,243	4.39	4,006
13 to 14	213,170	8.72	5,821
15 to 16	45,314	1.85	2,213
17 to 17	3,160	0.13	364
Frequency use of cigarette in the recent 30 days			
Never Smoked	1,995,785	81.68	27,741
None	312,587	12.79	7,126
1 or 2	72,910	2.98	3,458
3 to 5	23,340	0.96	1,651
6 to 9	11,636	0.48	1,304
10 to 19	9,292	0.38	898
20 or more	17,968	0.74	1,187
How did you acquired the cigarette in the recent 30 days			
Never Smoked	1,995,785	81.68	27,741
Did not Smoke	261,722	10.71	6,527
Store or Bar	47,411	1.94	2,566
Hawker	9,128	0.37	948

Table 4 continued from previous page

Variable	Frequency	%	Error
Someone bought it for me	17,727	0.73	1,235
Asked someone	36,846	1.51	1,939
Got it hidden from someone	31,953	1.31	1,715
Got it from na older person	16,290	0.67	1,108
Got it in another way	26,657	1.09	1,790
Someone refused to sell cigarettes			
Never Smoked	1,995,785	81.68	27,741
Did not try to buy	332,657	13.61	7,510
Yes, they refused to sell to me	38,270	1.57	2,241
No, I bought it	76,806	3.14	2,751

Source: Elaborated by the author using PeNSE data.

Drug-related variables (Table 5) demonstrate that 8.76% claim that has used some substance in life, with frequency pointing to a higher proportion of marijuana use (2.21%). Regarding sexual relation variables (Table 6), 27.07% of the sample declared that have had sex, with the number of sexual partners decreasing from one (9.97%) to five (1.43%), but increasing to 4.84% for six or more partners. To this subject, as PeNSE is based on a self-reported and anonymous questionnaire, students were free to answer the questions and this can be a source of bias. Consequently, likely the “6 or more sexual partners” category does not represent these student’s sexual lives. This is addressed as a limitation of our analysis. Finally, the use of preservatives was not reported for 8.26% of the 27.07% that declared to had sex, indicating a high proportion of unprotected sex.

TABLE 5 – Drug risk weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100.00	
Ever used drugs			
Yes	214,111	8.76	6,765
No	2,229,406	91.24	30,339
Age started to use drugs			
Never Used	2,229,406	91.24	30,339
7 to 8	5,266	0.22	665
9 to 10	3,827	0.16	440
11 to 12	30,371	1.24	1,775
13 to 14	132,125	5.41	4,799
15 to 16	40,216	1.65	2,093
17 to 18	2,307	0.09	371
Frequency use of drugs in the recent 30 days			
Never Used	2,229,406	91.24	30,339
None	112,782	4.62	4,346
1 or 2	50,962	2.09	2,557
3 to 5	19,627	0.80	1,463
6 to 9	10,078	0.41	1,074

Table 5 continued from previous page

Variable	Frequency	%	Error
10 or more	20,663	0.85	1,565
Frequency use of marijuana in the recent 30 days			
Never Used	2,229,406	91.24	30,339
None	114,381	4.68	4,393
1 or 2	53,956	2.21	2,544
3 to 9	21,089	0.86	1,476
10 or more	24,685	1.01	1,772
Frequency use of Crack in the recent 30 days			
Never Used	2,229,406	91.24	30,339
None	204,393	8.36	6,566
1 or 2	3,274	0.13	418
3 to 9	2,862	0.12	543
10 or more	3,581	0.15	485

Source: Elaborated by the author using PeNSE data.

TABLE 6 – Sexual risk weighted variables - PeNSE data

Variable	Frequency	%	Error
Total	2,443,517	100.00	
Ever had sexual intercourse			
Yes	661,393	27.07	13,332
No	1,782,124	72.93	26,672
Age sexual onset			
Never had it	1,782,124	72.93	26,672
9 to 10	66,049	2.70	2,736
11 to 12	101,032	4.13	3,884
13 to 14	371,924	15.22	7,891
15 to 16	113,558	4.65	3,615
17 to 18	8,830	0.36	820
Number of sexual partners			
Never had it	1,782,124	72.93	26,672
1	243,539	9.97	6,044
2	131,190	5.37	4,146
3	85,530	3.50	2,740
4	47,997	1.96	2,051
5	34,972	1.43	1,827
6 or more	118,165	4.84	3,913
Preservative use in the first sexual relation			
Never had it	1,782,124	72.93	26,672
Yes	402,959	16.49	8,366
No	258,434	10.58	7,367
Preservative use in the last sexual relation			
Never had it	1,782,124	72.93	26,672
Yes	438,385	17.94	8,925
No	201,818	8.26	6,143
Do not recall	21,190	0.87	1,430

Source: Elaborated by the author using PeNSE data.

In addition to this analysis, in the results section are presented descriptive tables by schooling expectation. It is worth noting that this data structure holds for chapters 1 and 2, as the same construction of data was implemented in the analysis. Moreover, the next section describes the method used to analyze these variables related to the schooling expectation.

1.3.2 Identification Strategy

This chapter aims to identify associations of features related to schooling expectation. With the exploratory purpose, the methodology implemented is ordered logistic regression. This pursuit relies on the large use of logit models in the literature, the nature of the study's data, and the exploratory purpose of the study. Such a model is a case of generalized linear models (GLM) applied to the categorical response context. As described by Agresti (2019), three components characterize a generalized linear model. First, a random component that identifies the response variable Y and assumes a probability distribution for it. Second, a linear predictor that specifies the explanatory variables, i.e., variables entering linearly on the equation's right hand. Third, a link function specifies how the expected value is related to the explanatory variables and connects the random component.

If the response variable is binary, the outcome can be classified as a success or failure. This is a case of standard logistic regression with link function equals to $g(\mu) = log[\mu/(1-\mu)]$, considering μ is structured as a probability between 0 and 1. But logistic regression generalizations apply to categorical responses with more than two categories labeled in an ordinal scale (AGRESTI, 2019). The most common ordinal logistic model is the proportional odds model, fitted with maximum likelihood estimator (MLE) based on a complex link function dependent on differences in logistic model probabilities (HARRELL, 2015). As well as in binary situations, quantitative and categorical data can be used as explanatory variables.

Therefore, the basic assumption for the ordinal regression model is that the response variable must behave in an ordinal fashion (HARRELL, 2015). This is the case for schooling expectations. Each education level is considered as a category that indicates a schooling prospect labeled on an ordinal scale. As they are: elementary, high school, technical high school, college, and other higher degrees. One of the components of a student's schooling expectations is "do not know". As this category has not an ordinal perspective, for these students, auxiliary binary regressions were carried out to analyze the main features related to not having an expectation.

Not all variables presented in the data subsection are used in the ordinal models. For socioeconomic and relational/behavioral variables, all the variables were maintained. For risk-related data, the models were divided into two, one considering the variables whether students have engaged in the risk, and the other using the frequency in the recent 30 days for alcohol, cigarette, drug variables, and the number of sexual partners for sex-related variables. Boards 1 and 2 address the variables used in the model, their

Board 1 – Variables used in the logistic model - Part 1

Response Variable		
Variable	**Type**	**Description**
Schooling Expectation	Categorical	5 scale variable representing elementary, high school,technical school, graduation and higher degrees.
Socioeconomic		
Gender	Binary	Female and males.
Age	Numeric	Age reported by the student.
Capital	Binary	Capital or interior.
Public	Binary	Public or private school.
Region	Categorical	North, northeast, midwest, southeast, south.
Color	Categorical	White, black, brown/mixed, yellow, indigenous.
People in the house	Numeric	Number of people living in the house.
Cellphone	Binary	Student has a cellphone.
Internet	Binary	Student has internet at home.
Car	Binary	Student's family has a car.
Housekeeper	Binary	Studen'ts family has a housemaid sometimes or always.
Lives w/ mother	Binary	Student lives with his mother.
Lives w/ father	Binary	Student lives with his father.
Relational/Behavioral		
Variable	**Type**	**Description**
Smoking Parents	Binary	At least one of the parents smoke.
Misses Class	Binary	Student missed the school without parents permissionin the recent 30 days.
Body	Categorical	Self-perception of student's weight. Obese, overweight, regular, thin, too thin.
Feels Lonely	Binary	Student have felt alone in the recent 12 months.
HPV	Binary	Student had HPV vaccine.
Parents understand problems	Binary	Parents understand the problems of their children.
Well-treated by colleagues	Binary	Colleagues treats the student kindly.
Victimizer	Binary	Student practice bullying in the recent 30 days.
Victimized	Binary	Student is a victim of bullying in the recent 30 days.
Bullying	Binary	Have the student already suffered bullying.
Close Friends	Numeric	Number of close friends.
Drug Friends	Numeric	If many friends of the student use drug.
Alcohol Friends	Numeric	If many friends of the student drink alcohol.
Rape or Assault	Binary	Student already suffered sexual violance.

Source: Elaborated by the author using PeNSE data.

type, and description.

Board 2 – Variables used in the logistic model - Part 2

Risk		
Variable	**Type**	**Description**
Alcohol	Binary	Have the student already drinked alcohol in his life.
Frequency alcohol	Categorical	Frequency of alcohol use in the recent 30 days.
Cigarette	Binary	Have the student already smoked cigarette.
Frequency cigarette	Categorical	Frequency of cigarette use in the recent 30 days.
Drug	Binary	Have the student already used drug.
Frequency Drug	Categorical	Frequency of drug use in the recent 30 days.
Sexual Relation	Binary	Have the student already involved in sexual intercourse.
Sexual partners	Categorical	How many people the student have been sexually related.
Relational/Behavioral		
Variable	**Type**	**Description**
Primary unit	Numeric	Primary sample unit controlling at school level.
Secondary unit	Numeric	Secondary sample unit controlling at class level.
Weight	Numeric	Sample weights.
Strata by region	Numeric	Stratified at region, federal unity, metropolitan region, capital level.

Source: Elaborated by the author using PeNSE data.

The separation of variables in the ordinal logistic regressions was verified by the data's nature, they are socioeconomic, relational/behavioral, risk-related. All models are controlled by gender and school type. Correlation analysis was implemented to address possible problems of multicollinearity. We verified that although socioeconomic and relational/behavioral variables (Figures 1 and 2) were not highly correlated, risk data variables (Figure 3) had a different pattern. As a precaution, many variables of this category were discarded for the model, as they were analyzed in the descriptive tables. The results are analyzed in the next section.

1.4 RESULTS

This section presents descriptive results by schooling expectation, the ordinal logistic regressions, and auxiliary binary logistic regressions. Technical high school and high school were joined to descriptive statistics, the same to the college and higher degrees. Table 7 returns descriptive socioeconomic results by schooling expectations. It can be verified there is a higher proportion of females with higher expectations (55.84&). Although males are 47.99% of the sex ratio for the sample data, they are 56.95% of those with elementary schooling prospects, 53.46% of high school/technical high school, and 50.39% of those who do not have an expectation.

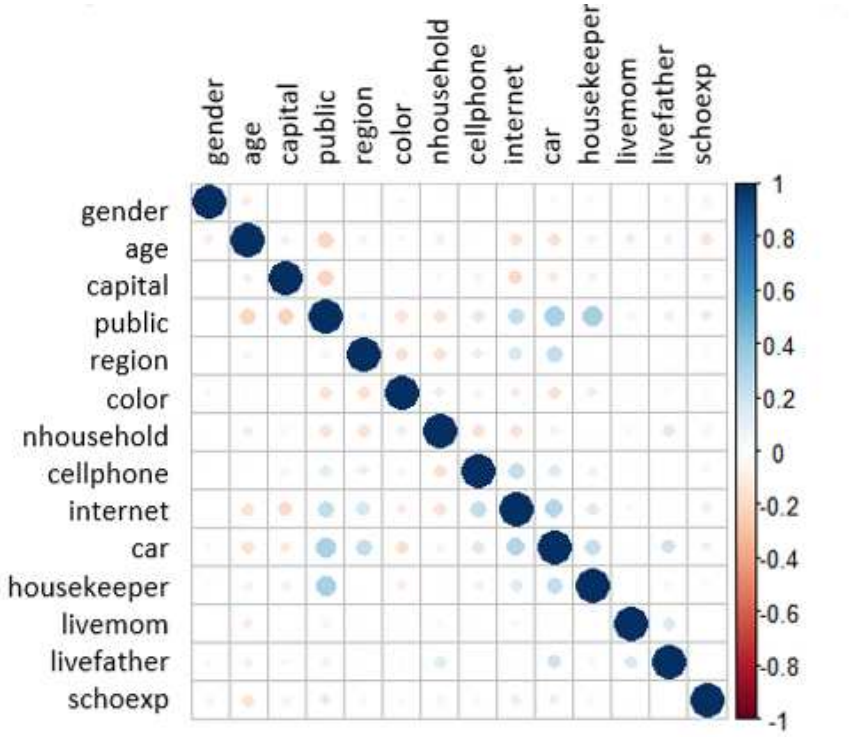

FIGURE 1 – Socioeconomic variables correlation

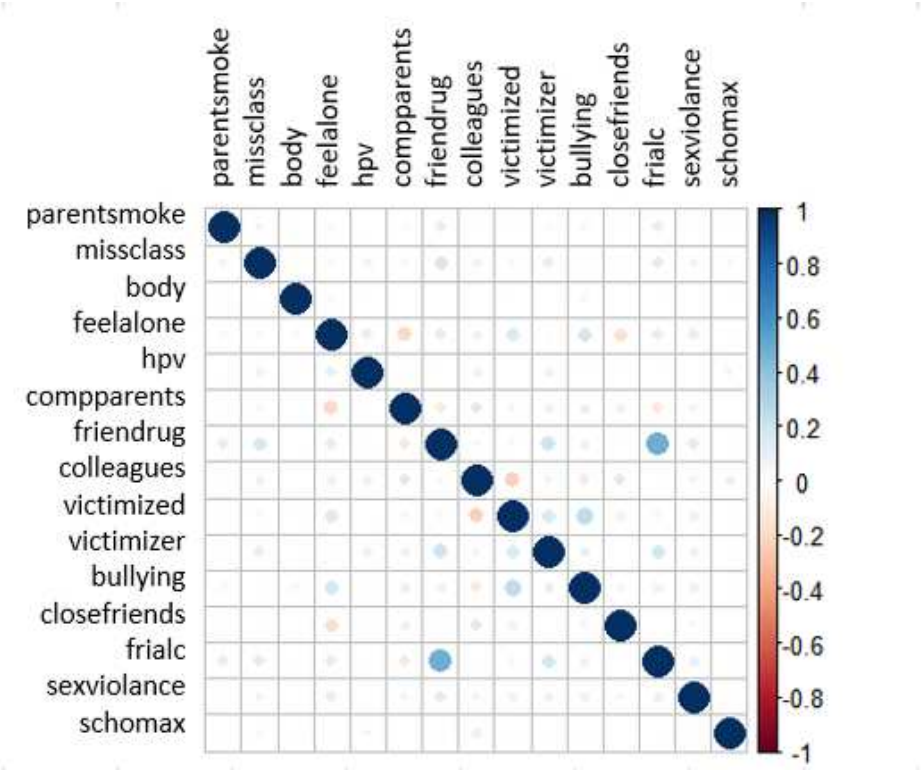

FIGURE 2 – Relational/behavioral variables correlation

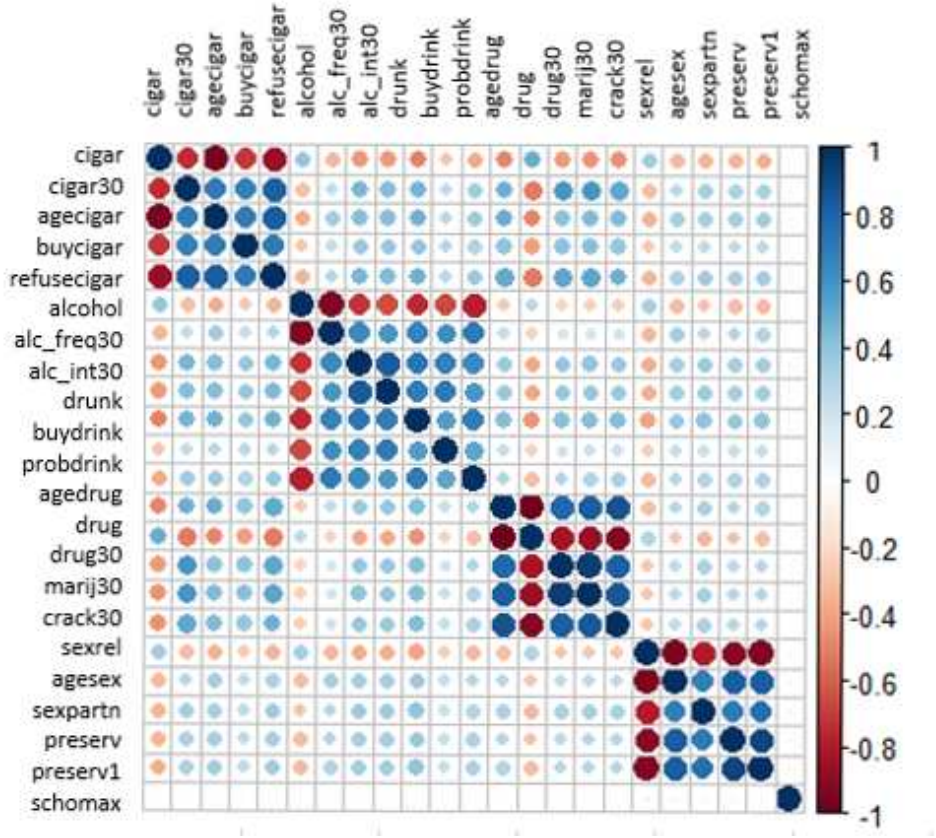

FIGURE 3 – Risk variables correlation

TABLE 7 – Socioeconomic characteristics and schooling expectations

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
Gender												
Girls	75,285	43.05	3,488	209,837	46.54	5,474	752,838	55.84	13,380	232,954	49.61	5,799
Boys	99,606	56.95	4,072	241,021	53.46	6,223	595,350	44.16	11,070	236,626	50.39	6,062
Age												
11	36	0.02	26	185	0.04	151	309	0.02	239	23	0.00	14
12	501	0.29	179	1,116	0.25	201	5,747	0.43	609	1,514	0.32	315
13	27,580	15.77	2,316	63,207	14.02	3,911	265,511	19.69	9,123	82,771	17.63	4,388
14	77,566	44.35	3,314	204,941	45.46	5,829	740,301	54.91	14,176	232,638	49.54	5,841
15	39,161	22.39	1,947	108,051	23.97	3,849	231,121	17.14	6,067	101,487	21.61	3,515
16	19,080	10.91	1,514	45,821	10.16	1,920	73,026	5.42	2,737	34,375	7.32	1,576
17	7,946	4.54	722	17,319	3.84	1,082	22,636	1.68	1,455	11,828	2.52	900
18	1,845	1.05	319	6,193	1.37	616	5,663	0.42	632	3,182	0.68	444
19	1,177	0.67	238	4,024	0.89	584	3,875	0.29	454	1,761	0.37	288
City												
Interior	33,232	19.00	1,532	85,624	18.99	2,845	341,699	25.35	6,703	100,895	21.49	3,161
Capital	141,660	81.00	5,715	365,234	81.01	9,251	1,006,489	74.65	20,412	368,685	78.51	9,137
Region												
North	12,847	7.35	646	45,938	10.19	1,901	133,041	9.87	3,971	43,191	9.20	1,566
Northeast	54,974	31.43	2,459	144,578	32.07	4,408	343,811	25.50	8,692	133,560	28.44	3,757
Southeast	79,573	45.50	5,137	189,473	42.03	8,012	591,337	43.86	17,980	201,681	42.95	8,256
South	15,652	8.95	1,317	49,233	10.92	2,375	166,175	12.33	6,058	56,670	12.07	2,623
Midwest	11,845	6.77	650	21,636	4.80	894	113,824	8.44	3,208	34,479	7.34	1,365
School type												
Private	12,727	7.28	1,513	36,812	8.16	3,527	259,032	19.21	17,766	54,054	11.51	4,012
Public	162,165	92.72	6,199	414,047	91.84	10,925	1,089,156	80.79	26,914	415,526	88.49	11,051
Skin Color												
White	55,014	31.46	2,699	139,895	31.03	4,550	527,855	39.15	13,383	163,280	34.77	4,905

Table 7 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Black	29,098	16.64	1,929	73,211	16.24	2,985	149,981	11.12	4,547	68,261	14.54	2,591
Yellow	7,631	4.36	850	17,415	3.86	1,260	54,988	4.08	2,226	20,500	4.37	1,276
Brown/mixed-Skin	77,348	44.23	3,343	203,100	45.05	5,696	574,595	42.62	11,652	201,229	42.85	5,712
Indígenous	5,800	3.32	623	17,237	3.82	1,317	40,768	3.02	1,959	16,309	3.47	1,153
Has a Cellphone												
Yes	148,513	84.92	5,399	384,693	85.32	8,533	1,208,067	89.61	20,239	400,572	85.30	8,627
No	26,378	15.08	1,795	66,165	14.68	2,797	140,121	10.39	3,854	69,008	14.70	2,790
Internet												
Yes	122,201	69.87	4,823	314,278	69.71	7,670	1,124,582	83.41	19,978	344,963	73.46	8,246
No	52,690	30.13	2,661	136,580	30.29	4,597	223,606	16.59	5,594	124,617	26.54	3,692
Has a Car												
Yes	85,701	49.00	3,736	218,867	48.54	5,526	828,454	61.45	16,872	252,322	53.73	6,887
No	89,190	51.00	3,686	231,991	51.46	6,189	519,734	38.55	11,074	217,257	46.27	5,494
Housekeeper												
Yes	16,690	9.54	1,272	27,932	6.20	1,661	148,388	11.01	7,113	37,045	7.89	1,927
No	158,201	90.46	5,754	422,926	93.80	9,467	1,199,800	88.99	20,297	432,535	92.11	9,494
Lives with the Mother												
Yes	156,240	89.34	5,533	402,685	89.32	8,952	1,221,150	90.58	20,356	420,198	89.48	8,943
No	18,652	10.66	1,373	48,173	10.68	2,104	127,038	9.42	3,344	49,381	10.52	1,971
Lives with the Father												
Yes	109,323	62.51	4,170	286,752	63.60	6,631	862,328	63.96	15,431	296,479	63.14	7,142
No	65,569	37.49	3,178	164,106	36.40	4,976	485,860	36.04	9,226	173,101	36.86	4,627
Number of People in the Residence												
1	1,099	0.63	409	1,118	0.25	277	1,181	0.09	340	682	0.15	284
2	9,182	5.25	942	21,696	4.81	1,322	80,381	5.96	3,012	25,033	5.33	1,397
3	28,643	16.38	1,616	81,702	18.12	2,737	276,867	20.54	6,275	82,883	17.65	2,998
4	49,985	28.58	2,211	134,611	29.86	4,002	481,412	35.71	9,554	151,488	32.26	4,046
5	40,482	23.15	2,206	98,055	21.75	3,564	284,063	21.07	6,484	104,705	22.30	3,569
6	22,771	13.02	1,675	57,348	12.72	2,692	123,460	9.16	4,173	50,996	10.86	2,495

Table 7 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
7	9,348	5.35	1,027	25,263	5.60	1,448	51,364	3.81	2,169	24,925	5.31	1,597
8 or more	13,381	7.65	1,074	31,065	6.89	1,869	49,459	3.67	1,980	28,868	6.15	1,786

Source: Elaborated by the author using PeNSE data.

Regarding their age, it is possible to see for 14 years-old students, a higher proportion is found to have college/higher expectations. Nevertheless, this dynamic is altered from 15 to 19 years-old, with elementary and high school/high technical school becoming more prevalent. Also, in urban cities, the proportion of those expecting college or more was lower (74.65%) than the other expectations (81.00%; 78.51%). Those aspects can be related to job prospection in urban cities, which might instigate students to leave school earlier.

The type of school (private or public) also presents differences in the proportions. For private schools, the proportion of college or higher degrees (19.21%) is above the one found for the whole sample (14.84%), while students from public schools are overrepresented in the lower expectations or “do not know” case. Related to their color, white and black present more variance. Compared to the national proportion of high expectation (white=36.26% and black=13.12%), white students are overrepresented with 39.15% of high expectation, while black students are 11.12%.

Moreover, compared to the national proportions, those with higher expectations have more cellphones (89.61%), internet at home (83.41%), car (61.45%), and a housekeeper (11.01%). Living with parents did not present many variations, but more people in the house contributes to a lower proportion of high expectations compared to the national sample.

Table 8 has descriptive relational/behavioral statistics by schooling expectation. It is possible to identify that a lower proportion of parents who smoke (24.40%), of students who miss classes without parents’ permission (20.17%), and of students practicing (18.62%) or suffering (22.47%) victimization in the recent 30 days. At the same time, there is a higher proportion of students feeling loneliness (18.21%), of taking the HPV vaccine (43.01%), have parents understanding their problems (45.08%), and being well-treated by their colleagues (68.40%). The other variables have not presented much difference related to the national proportions.

TABLE 8 – Relational, behavioral characteristics and schooling expectations

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
At least one of the parents smoke												
Yes	53,693	30.70	2,623	127,403	28.26	4,434	328,964	24.40	7,280	127,885	27.23	4,486
No	121,198	69.30	4,480	323,455	71.74	7,258	1,019,224	75.60	17,519	341,695	72.77	7,100
Missed class without parental permission												
Yes	55,618	31.80	2,794	123,091	27.30	4,698	271,973	20.17	7,138	115,690	24.64	4,222
No	119,273	68.20	4,228	327,767	72.70	7,324	1,076,215	79.83	17,707	353,890	75.36	7,128
Body Perception												
Very Thin	11,534	6.59	1,071	25,084	5.56	1,515	65,493	4.86	2,503	26,605	5.67	1,660
Thin	32,995	18.87	1,929	93,225	20.68	3,466	278,952	20.69	6,907	93,554	19.92	3,072
Normal	104,719	59.88	4,117	267,574	59.35	6,390	728,470	54.03	12,182	264,710	56.37	6,113
Fat	21,417	12.25	1,754	59,011	13.09	2,628	243,046	18.03	5,734	75,015	15.97	2,849
Very Fat	4,227	2.42	667	5,964	1.32	620	32,228	2.39	1,771	9,696	2.06	1,040
Feels Lonely												
Yes	23,007	13.15	1,645	61,124	13.56	2,542	245,516	18.21	5,803	75,737	16.13	2,793
No	151,885	86.85	5,273	389,734	86.44	8,870	1,102,672	81.79	18,083	393,842	83.87	8,345
HPV Vaccination												
Yes	53,064	30.34	2,764	145,840	32.35	4,555	579,836	43.01	11,654	167,391	35.65	4,903
No	121,828	69.66	4,481	305,018	67.65	7,082	768,352	56.99	12,993	302,189	64.35	6,969
Parents Understand their Problems												
Yes	78,477	44.87	3,670	194,335	43.10	5,296	607,789	45.08	11,368	191,554	40.79	4,888
No	96,415	55.13	3,563	256,523	56.90	6,172	740,399	54.92	13,314	278,026	59.21	6,672
Well-Treated by the Colleagues												
Yes	88,016	50.33	3,568	247,915	54.99	6,211	922,142	68.40	15,745	271,208	57.76	6,073
No	86,875	49.67	3,489	202,943	45.01	5,366	426,046	31.60	8,929	198,372	42.24	5,190
Practiced Vitimization												
Yes	42,188	24.12	2,582	92,725	20.57	3,488	251,006	18.62	6,614	93,964	20.01	3,652
No	132,703	75.88	4,630	358,133	79.43	8,025	1,097,182	81.38	17,875	375,615	79.99	7,961

Table 8 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Suffered Victimization												
Yes	50,146	28.67	2,640	114,976	25.50	3,949	302,927	22.47	6,544	109,702	23.36	3,635
No	124,746	71.33	4,608	335,882	74.50	7,448	1,045,261	77.53	17,717	359,878	76.64	7,783
Suffered Bullying					0.00			0.00				
Yes	78,480	44.87	3,676	198,099	43.94	5,801	676,211	50.16	11,809	212,604	45.28	5,595
No	96,412	55.13	3,778	252,759	56.06	6,162	671,977	49.84	12,396	256,976	54.72	5,945
Number of Close Friends												
None	8,699	4.97	978	20,187	4.48	1,505	48,771	3.62	2,126	24,222	5.16	1,573
1	12,204	6.98	1,106	29,759	6.60	1,644	81,425	6.04	3,037	28,676	6.11	1,544
2	18,844	10.77	1,364	54,174	12.02	2,804	182,079	13.51	5,077	53,901	11.48	2,454
3 or more	135,144	77.27	4,873	346,739	76.91	7,877	1,035,913	76.84	16,924	362,781	77.26	7,465
Friends who drink alcohol												
None or don't know	50,193	28.70	2,565	124,615	27.64	4,467	356,972	26.48	8,755	149,343	31.80	4,745
Few	49,161	28.11	2,489	120,419	26.71	3,503	389,365	28.88	7,702	123,495	26.30	3,554
Some	35,844	20.49	1,900	109,476	24.28	3,816	329,608	24.45	6,992	102,033	21.73	3,349
Most of	31,933	18.26	1,965	80,200	17.79	3,065	239,485	17.76	5,904	81,318	17.32	2,738
All	7,761	4.44	1,002	16,149	3.58	1,255	32,759	2.43	1,970	13,391	2.85	1,110
Friends who use Drugs												
None or don't know	105,245	60.18	3,727	282,338	62.62	6,621	843,616	62.57	15,648	300,582	64.01	6,394
Few	32,188	18.40	2,231	82,541	18.31	3,330	282,231	20.93	7,267	85,126	18.13	3,508
Some	21,193	12.12	1,539	56,065	12.44	2,514	158,522	11.76	4,718	54,655	11.64	2,686
Most of	12,361	7.07	1,403	25,407	5.64	1,766	57,381	4.26	3,225	24,919	5.31	1,982
All	3,904	2.23	880	4,507	1.00	705	6,438	0.48	715	4,298	0.92	679
Victim of sexual violence												
Yes	8,549	4.89	759	19,481	4.32	1,484	47,095	3.49	2,232	18,328	3.90	1,272
No	166,343	95.11	5,717	431,377	95.68	9,271	1,301,093	96.51	21,067	451,252	96.10	9,374

Source: Elaborated by the author using PeNSE data.

Concerning the alcohol risk data, Table 9 demonstrates that students with lower expectations had a diminished proportion of alcohol experimentation than the national sample. On the other hand, those with elementary schooling prospects, who had experimented alcohol had a higher frequency than the national sample and the other expectations. As an example, the “20 or more days” frequency category (2.58%), “4 or more glasses” intensity (7.17%), “10 or more times that got drunk” (4.27%), and the number of times had problems due to alcohol drinking (1.94%), compared to, respectively, 1.24%, 6.50%, 2.43%, 1.00% in the national sample.

TABLE 9 – Risk Descriptive Statistics (alcohol) and Schooling Expectation

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
Drank alcohol at least once												
Yes	91,134	52.11	3,931	246,186	54.60	6,437	721,942	53.55	14,183	257,427	54.82	6,417
No	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
Age started to drink alcohol												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
7 to 8	3,696	2.11	499	10,881	2.41	1,043	39,005	2.89	2,043	13,244	2.82	940
9 to 10	7,313	4.18	816	18,959	4.21	1,451	60,529	4.49	3,171	23,815	5.07	1,631
11 to 12	19,818	11.33	1,623	54,267	12.04	2,669	179,939	13.35	4,987	60,295	12.84	2,987
13 to 14	44,514	25.45	2,439	122,614	27.20	3,959	380,277	28.21	8,652	130,377	27.76	4,040
15 to 16	14,055	8.04	1,209	35,373	7.85	1,736	57,728	4.28	2,122	27,386	5.83	1,386
17 to 18	1,738	0.99	291	4,091	0.91	452	4,463	0.33	497	2,309	0.49	379
Frequency drank alcohol in the recent 30 days												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
None	43,854	25.08	2,576	130,432	28.93	4,126	417,030	30.93	9,024	139,917	29.80	4,022
1 or 2 days	27,290	15.60	1,774	70,536	15.64	2,636	182,134	13.51	4,571	73,980	15.75	3,020
3 to 5 days	7,179	4.11	867	18,778	4.16	1,313	56,379	4.18	2,341	18,524	3.94	1,533
6 to 9 days	5,090	2.91	695	11,044	2.45	1,077	30,658	2.27	1,729	10,656	2.27	934
10 to 19 days	3,209	1.84	535	8,188	1.82	974	23,402	1.74	1,768	8,024	1.71	835
20 or more days	4,512	2.58	841	7,208	1.60	749	12,340	0.92	910	6,327	1.35	698
Intensity of alcohol use in the recent 30 days												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
Did not drink	38,797	22.18	2,061	117,638	26.09	3,834	383,008	28.41	8,489	129,801	27.64	3,847
1 Glass or less	28,324	16.20	2,015	70,026	15.53	2,803	169,739	12.59	4,955	69,085	14.71	2,851
2 to 3 glasses	11,473	6.56	963	29,854	6.62	1,900	82,183	6.10	2,789	27,817	5.92	1,912
4 or more glasses	12,541	7.17	1,227	28,668	6.36	1,518	87,012	6.45	3,366	30,724	6.54	1,810
Number of times got really Drunk												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422

Table 9 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
None	46,153	26.39	2,337	143,544	31.84	4,273	445,884	33.07	9,530	156,670	33.36	4,729
1 to 2	27,005	15.44	1,967	67,958	15.07	2,870	181,538	13.47	4,908	64,388	13.71	2,310
3 to 5	8,076	4.62	922	18,617	4.13	1,357	49,943	3.70	2,385	19,389	4.13	1,372
6 to 9	2,440	1.39	443	4,875	1.08	799	15,395	1.14	1,343	5,555	1.18	867
10 or more	7,461	4.27	1,044	11,192	2.48	1,029	29,182	2.16	1,669	11,426	2.43	1,300
How did you acquired acohol in the recent 30 days												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
Did not drink	32,299	18.47	1,940	100,470	22.28	3,463	330,182	24.49	7,629	108,681	23.14	3,404
Store or Bar	9,334	5.34	1,027	24,352	5.40	1,586	51,897	3.85	2,394	20,587	4.38	1,518
Hawker	2,507	1.43	533	2,956	0.66	489	4,933	0.37	609	1,765	0.38	277
Someone bought it for me	3,530	2.02	669	5,001	1.11	719	12,669	0.94	1,302	5,564	1.18	788
Friends	11,360	6.50	1,050	28,587	6.34	1,691	69,163	5.13	2,636	25,305	5.39	1,550
Got it hidden from someone	1,842	1.05	380	3,616	0.80	533	15,631	1.16	1,203	6,986	1.49	815
From a person older than me	3,757	2.15	467	12,821	2.84	1,000	36,747	2.73	1,759	16,162	3.44	1,559
Party	23,275	13.31	1,770	61,372	13.61	2,565	180,122	13.36	4,991	63,179	13.45	2,563
Got it from another way	3,230	1.85	726	7,010	1.55	950	20,598	1.53	1,339	9,199	1.96	1,053
Times that had problems due to alcohol drinking												
Never drank	83,757	47.89	3,548	204,673	45.40	5,401	626,246	46.45	12,223	212,153	45.18	5,422
None	73,775	42.18	3,253	212,128	47.05	5,779	635,900	47.17	12,762	224,512	47.81	5,772
1 or 2 ever	9,826	5.62	1,120	21,635	4.80	1,568	54,766	4.06	2,395	19,733	4.20	1,297
3 to 5 ever	3,050	1.74	578	5,669	1.26	711	14,985	1.11	1,144	6,168	1.31	791
6 to 9 ever	1,083	0.62	255	1,633	0.36	392	5,039	0.37	640	2,246	0.48	442
10 or more ever	3,399	1.94	615	5121	1.14	698	11,252	0.83	873	4,767	1.02	623

Source: Elaborated by the author using PeNSE data.

For cigarette use (Table 10), a lower proportion was found in students with college/higher degrees (17.32%) expectations compared to the national sample (18.32%). The frequency in use was higher for learners with the elementary prospect. They were also more able to buy cigarettes with no refusion (4.13%), compared to 2.74% of college prospect students and 3.14% of the national sample. Therefore, although alcohol experimentation was more seen in high-level expectation students, the frequency used was verified more for youth with low expectations. Nevertheless, the cigarette case had a reversal to the experimentation case, with a higher proportion to low prospect students, but a higher frequency in use was seen at this stratum.

TABLE 10 – Risk Descriptive Statistics (cigarette) and Schooling Expectation

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
Smoked Cigarette Once												
Yes	33,709	19.27	2,069	87,871	19.49	3,412	233,559	17.32	6,297	92,593	19.72	3,385
No	141,182	80.73	4,878	362,987	80.51	8,186	1,114,629	82.68	18,741	376,986	80.28	7,937
Age started to smoke												
Never Smoked	141,182	80.73	4,878	362,987	80.51	8,186	1,114,629	82.68	18,741	376,986	80.28	7,937
7 to 8	2,612	1.49	607	6,713	1.49	793	18,647	1.38	1,151	8,898	1.89	998
9 to 10	2,897	1.66	482	9,077	2.01	876	20,080	1.49	1,247	9,920	2.11	1,013
11 to 12	8,167	4.67	965	20,449	4.54	1,569	56,674	4.20	2,708	21,953	4.68	1,528
13 to 14	15,003	8.58	1,259	38,845	8.62	2,109	116,557	8.65	3,700	42,765	9.11	2,074
15 to 16	4,599	2.63	703	11,641	2.58	1,054	20,567	1.53	1,275	8,507	1.81	790
17 to 17	431	0.25	133	1,145	0.25	232	1,034	0.08	198	551	0.12	148
Frequency use of cigarette in the recent 30 days												
Never Smoked	141,182	80.73	4,878	362,987	80.51	8,186	1,114,629	82.68	18,741	376,986	80.28	7,937
None	20,021	11.45	1,408	57,057	12.66	2,347	170,700	12.66	4,634	64,809	13.80	2,579
1 or 2	6,334	3.62	774	18,420	4.09	1,582	34,211	2.54	2,213	13,946	2.97	1,114
3 to 5	3,528	2.02	795	4,021	0.89	644	10,188	0.76	953	5,603	1.19	841
6 to 9	664	0.38	203	2,838	0.63	561	5,284	0.39	766	2,850	0.61	608
10 to 19	1,247	0.71	359	1,442	0.32	265	5,328	0.40	741	1,275	0.27	241
20 or more	1,916	1.10	431	4,093	0.91	517	7,849	0.58	768	4,111	0.88	559
How did you acquired the cigarette in the recent 30 days												
Never Smoked	141,182	80.73	4,878	362,987	80.51	8,186	1,114,629	82.68	18,741	376,986	80.28	7,937
Did not Smoke	15,985	9.14	1,304	46,738	10.37	2,078	144,571	10.72	4,117	54,428	11.59	2,385
Store or Bar	4,905	2.80	637	11,454	2.54	1,162	22,008	1.63	1,429	9,045	1.93	1,052
Hawker	912	0.52	359	1,973	0.44	413	4,337	0.32	527	1,907	0.41	456
Someone bought it for me	1,678	0.96	357	4,530	1.00	603	8,088	0.60	908	3,431	0.73	526
Asked someone	3,063	1.75	539	7,538	1.67	816	19,469	1.44	1,512	6,776	1.44	634
Got it hidden from someone	2,928	1.67	794	7,893	1.75	817	15,129	1.12	1,171	6,002	1.28	691

Table 10 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Got it from na older person	2,004	1.15	367	3,213	0.71	475	7,119	0.53	716	3,953	0.84	675
Got it in another way	2,234	1.28	496	4,532	1.01	731	12,838	0.95	1,105	7,052	1.50	966
Somene refused to sell cigarettes												
Never Smoked	141,182	80.73	4,878	362,987	80.51	8,186	1,114,629	82.68	18,741	376,986	80.28	7,937
Did not try to buy	22,903	13.10	1,612	60,315	13.38	2,563	181,421	13.46	4,984	68,019	14.49	2,692
Yes, they refused to sell to me	3,576	2.04	500	10,325	2.29	1,203	15,148	1.12	1,158	9,221	1.96	1,095
No, I bought it	7,230	4.13	871	17,232	3.82	1,368	36,990	2.74	1,804	15,354	3.27	1,220

Source: Elaborated by the author using PeNSE data.

Related to drug, descriptive statistics are presented in Table 11. It is possible to identify that students with elementary prospects have a higher proportion of drug experimentation compared to the national sample and other expectations/not expectations. Further, though a high-level use represents a minor proportion for all groups, students with elementary expectations demonstrated a higher use proportion for “10 or more days” for almost all frequency drug variables related to other groups, except marijuana use for those that do not know their expectations.

TABLE 11 – Risk Descriptive Statistics (drugs) and Schooling Expectation

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
Ever used drugs												
Yes	19,560	11.18	1,822	39,777	8.82	2,503	113,609	8.43	4,216	41,165	8.77	2,496
No	155,331	88.82	5,221	411,081	91.18	8,809	1,234,579	91.57	19,774	428,415	91.23	8,659
Age started to use drugs												
Never Used	155,331	88.82	5,221	411,081	91.18	8,809	1,234,579	91.57	19,774	428,415	91.23	8,659
7 to 8	612	0.35	265	1,434	0.32	391	1,972	0.15	330	1,248	0.27	345
9 to 10	668	0.38	272	786	0.17	197	1,439	0.11	221	934	0.20	176
11 to 12	2,519	1.44	571	4,950	1.10	738	16,802	1.25	1,415	6,101	1.30	803
13 to 14	10,426	5.96	1,061	22,881	5.07	1,912	73,351	5.44	3,149	25,467	5.42	1,986
15 to 16	4,843	2.77	978	9,158	2.03	809	19,059	1.41	1,279	7,156	1.52	764
17 to 18	492	0.28	190	569	0.13	137	986	0.07	241	259	0.06	95
Frequency use of drugs in the recent 30 days												
Never Used	155,331	88.82	5,221	411,081	91.18	8,809	1,234,579	91.57	19,774	428,415	91.23	8,659
None	9,283	5.31	1,069	18,799	4.17	1,326	63,877	4.74	3,001	20,822	4.43	1,697
1 or 2	4,794	2.74	736	11,158	2.47	1,368	25,788	1.91	1,635	9,221	1.96	952
3 to 5	1,765	1.01	630	3,803	0.84	653	9,986	0.74	1,027	4,072	0.87	625
6 to 9	567	0.32	175	2,277	0.50	485	4,832	0.36	669	2,402	0.51	623
10 or more	3,151	1.80	745	3,740	0.83	588	9,125	0.68	937	4,647	0.99	756
Frequency use of marijuana in the recent 30 days												
Never Used	155,331	88.82	5,221	411,081	91.18	8,809	1,234,579	91.57	19,774	428,415	91.23	8,659
None	9,933	5.68	1,290	20,001	4.44	1,467	65,391	4.85	3,110	19,056	4.06	1,351
1 or 2	5,061	2.89	753	11,606	2.57	1,308	25,360	1.88	1,445	11,929	2.54	1,129
3 to 9	2,475	1.41	712	3,730	0.83	664	11,155	0.83	1,070	3,729	0.79	595
10 or more	2,092	1.20	465	4,440	0.98	633	11,703	0.87	1,070	6,451	1.37	830
Frequency use of Crack in the recent 30 days												
Never Used	155,331	88.82	5,221	411,081	91.18	8,809	1,234,579	91.57	19,774	428,415	91.23	8,659
None	17,868	10.22	1,682	37,707	8.36	2,452	109,654	8.13	4,121	39,165	8.34	2,484

Table 11 continued from previous page

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
1 or 2	475	0.27	159	712	0.16	171	1,282	0.10	264	805	0.17	232
3 to 9	677	0.39	298	627	0.14	165	1,116	0.08	343	443	0.09	254
10 or more	541	0.31	256	732	0.16	232	1,556	0.12	273	753	0.16	194

Source: Elaborated by the author using PeNSE data.

Lastly, Table 12 presents sexual statistics by schooling expectation. As the expectation becomes higher, the proportion of students reporting ever had sexual intercourse becomes smaller, with those that do not have an expectation being close to the national sample proportion. The same pattern occurs to other sex-related variables presented in the table. However, special attention is drawn to preservative use among students. A high proportion of the students who claimed to have sexual intercourse reported not using the preservative in their relations. Even with elementary prospect students having higher proportions, not using condoms was quite large for all groups. Although other contraceptive methods are not brought to this study, condoms represent the main fashion of contraception to pregnancy and sexually transmitted diseases (STDs). In such a context, these proportions are a cause of concern.

TABLE 12 – Risk Descriptive Statistics (Sexual Relations) and Schooling Expectation

	Elementary			High School/Technical School			College/Higher degrees			Do not know		
	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error	Frequency	%	Error
Total	174,891			450,858			1,348,188			469,580		
Ever had sexual intercourse												
Yes	60,982	34.87	2,950	143,216	31.77	4,556	329,887	24.47	7,119	127,308	27.11	3,846
No	113,909	65.13	4,427	307,642	68.23	7,305	1,018,301	75.53	18,100	342,272	72.89	7,782
Age that had the first sexual relation												
Never had it	113,909	65.13	4,427	307,642	68.23	7,305	1,018,301	75.53	18,100	342,272	72.89	7,782
9 to 10	7,274	4.16	949	14,316	3.18	1,095	31,320	2.32	1,958	13,140	2.80	1,054
11 to 12	9,968	5.70	1,068	21,880	4.85	1,502	49,754	3.69	2,287	19,430	4.14	1,430
13 to 14	31,217	17.85	2,061	76,354	16.94	2,969	193,068	14.32	4,627	71,286	15.18	2,660
15 to 16	11,210	6.41	972	27,779	6.16	1,409	52,314	3.88	2,102	22,256	4.74	1,350
17 to 18	1,313	0.75	259	2,888	0.64	370	3,432	0.25	492	1,197	0.25	225
Number of sexual partners												
Never had it	113,909	65.13	4,427	307,642	68.23	7,305	1,018,301	75.53	18,100	342,272	72.89	7,782
1	18,303	10.47	1,266	48,150	10.68	2,216	131,919	9.78	4,191	45,168	9.62	1,883
2	13,856	7.92	1,137	30,253	6.71	1,843	63,146	4.68	2,417	23,936	5.10	1,392
3	7,332	4.19	774	19,087	4.23	1,198	40,163	2.98	1,781	18,948	4.04	1,233
4	4,847	2.77	662	9,228	2.05	827	23,052	1.71	1,294	10,871	2.31	1,104
5	3,298	1.89	672	9,257	2.05	1,021	15,963	1.18	1,123	6,454	1.37	754
6 or more	13,347	7.63	1,199	27,242	6.04	1,676	55,644	4.13	2,081	21,932	4.67	1,497
Used preservative in the first sexual relation												
Never had it	113,909	65.13	4,427	307,642	68.23	7,305	1,018,301	75.53	18,100	342,272	72.89	7,782
Yes	36,485	20.86	2,174	85,614	18.99	2,891	206,149	15.29	4,905	74,711	15.91	2,775
No	24,497	14.01	1,683	57,602	12.78	3,034	123,738	9.18	4,096	52,597	11.20	2,267
Used preservative in the last sexual relation												
Never had it	113,909	65.13	4,427	307,642	68.23	7,305	1,018,301	75.53	18,100	342,272	72.89	7,782
Yes	40,999	23.44	2,118	94,011	20.85	2,994	222,024	16.47	5,383	81,351	17.32	2,888
No	17,621	10.08	1,478	44,706	9.92	2,599	99,649	7.39	3,422	39,843	8.48	2,005
Do not recall	2,362	1.35	512	4,499	1.00	577	8,215	0.61	931	6,114	1.30	660

Source: Elaborated by the author using PeNSE data.

Analyzing ordinal logistic regressions, Table 13 returns the results for socioeconomic variables. The coefficients analysis considers holding all other variables constant (*ceteris paribus*). In this situation, for male students, the odds of having high schooling expectations are 38.82% lower than for female students. Rural students are 13.60% less likely to have high expectations compared to urban students. Public school is associated with a 44.96% decrease in having high expectations related to private schools.

Further, black students are 22.80% more likely to have a lower expectation than white peers, and students from the north region are more likely to have high expectations compared to other regions. Revolving variables related to economic possessions, the odds of being more likely to have high expectations are 1.59 times for students with the internet at home. The same pattern follows having a cellphone (OR=1.09) and a car in the family (OR=1.26), with lower magnitudes. So, such results demonstrate that internet access is an important feature for youth because the impact on school achievement is significant. Also, the odds of having a high expectation are 1.10 times for those with a housekeeper. Once we do not control for income variables, internet access, housekeeper, and the other variables mentioned can be considered a proxy of the economic situation. Hence, they demonstrated that better economic conditions are aligned with higher expectations.

Moreover, the odds of being more likely to have higher expectations is 1.11 for those living with their mothers. On the other hand, living with the father did not demonstrate to be a significant variable to schooling expectation. Also, each person in the house decreases the odds of having high expectations by 9.06%. The model intercepts are all significant; this means that the expectation groups are different regarding their socioeconomic variables. Besides, the intercept becomes higher in magnitude as the expected increases. It means that, on average, the start point for those with higher expectations is also high in their socioeconomic status.

Table 14 has the ordinal logistic regression results for relational/behavioral variables. For students whose parents smoke, the odds of having high expectations are 12.56% lower. Missing classes without parents' permission reduces the odds of having high expectations by 26.19%. Also, students with a body perception of being 'very thin' are less likely to have high expectations than all other body perception categories. Some other features related to higher odds of having a bigger expectation are HPV vaccination (OR=1.33), parents that understand students' problems (OR=1.07) and being well-treated by colleagues (OR=1.63).

On the other hand, some results might appear counterintuitive. For instance, feeling lonely in the recent 12 months is associated with 1.37 times the odds of having high expectations. Also, students who suffered victimization in the recent 30 days are 15.32% less likely to have high expectations, alternatively being a victim of bullying is associated with 27.00% the odds of being more likely to have a high expectation. In this context, it seems like feelings of belonging may influence students' perspectives, and this could be highly altered during adolescence transition.

TABLE 13 – Ordinal logistic regression - Socioeconomic variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Gender				
Female	1.00			
Male	0.61	0.58	0.65	<0.01
City				
Capital	1.00			
Interior	0.86	0.81	0.92	<0.01
Region				
North	1.00			
Northeast	0.77	0.71	0.83	<0.01
Southeast	0.74	0.68	0.80	<0.01
South	0.87	0.78	0.96	<0.01
Midwest	1.05	0.96	1.14	0.279
School Type				
Private	1.00			
Public	0.55	0.50	0.61	<0.01
Skin Color				
White	1.00			
Black	0.77	0.70	0.85	<0.01
Yellow	0.90	0.80	1.02	0.099
Brown/mixed-Skin	0.95	0.90	1.01	0.121
Indígenous	0.87	0.76	1.00	0.058
Has a Cellphone				
No	1.00			
Yes	1.09	1.00	1.19	0.046
Internet				
No	1.00			
Yes	1.59	1.50	1.70	<0.01
Has a Car				
No	1.00			
Yes	1.26	1.19	1.33	<0.01
Housekeeper				
No	1.00			
Yes	1.11	1.01	1.23	0.025
Lives with the Mother				
No	1.00			
Yes	1.11	1.03	1.20	<0.01
Lives with the Father				
No	1.00			
Yes	0.98	0.93	1.03	0.450
People in the house	0.91	0.90	0.92	<0.01
Intercepts				
Elem|High	0.04	0.03	0.04	<0.01
High|HigTec	0.12	0.12	0.13	<0.01
HigTec|Coll	0.18	0.17	0.19	<0.01
Coll|PosGrad	0.41	0.40	0.43	<0.01

Source: Elaborated by the author using PeNSE data.

TABLE 14 – Ordinal Logistic Regression - Relational/Behavioral variables

Variables	Schooling Expectations			
	OR	IC 95%		p-value
		Inferior	Superior	
Parents smoke				
No	1.00			
Yes	0.87	0.82	0.93	<0.01
Missed classes				
No	1.00			
Yes	0.74	0.69	0.79	<0.01
Body Perception				
Very Thin	1.00			
Thin	1.21	1.07	1.37	<0.01
Normal	1.10	0.99	1.23	0.075
Fat	1.50	1.32	1.70	<0.01
Very Fat	1.46	1.18	1.81	<0.01
Feels Lonely				
No	1.00			
Yes	1.37	1.26	1.50	<0.01
HPV Vaccination				
No	1.00			
Yes	1.33	1.22	1.46	<0.01
Comprehensive parents				
No	1.00			
Yes	1.07	1.02	1.12	<0.01
Well-Treated by the Colleagues				
No	1.00			
Yes	1.63	1.54	1.72	<0.01
Practiced Vitimization				
No	1.00			
Yes	0.95	0.89	1.02	0.145
Suffered Victimization				
No	1.00			
Yes	0.85	0.79	0.91	<0.01
Suffered Bullying				
No	1.00			
Yes	1.27	1.20	1.34	<0.01
Close Friends	1.01	0.98	1.05	0.461
Alcohol Friends	1.04	1.01	1.07	0.011
Drug Friends	0.95	0.91	0.99	0.011
Victim of sexual violence				
No	1.00			
Yes	0.93	0.82	1.06	0.297
Intercepts				
Elem|High	0.08	0.07	0.10	<0.01
High|HigTec	0.28	0.27	0.30	<0.01
HigTec|Coll	0.43	0.40	0.45	<0.01
Coll|PosGrad	0.96	0.93	0.99	0.687

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school

Still considering Table 14, having friends who drink alcohol is associated with 1.04 the odds of having high expectations, but for friends who use drugs, the odds decrease 5% compared to those who do not have friends consuming these goods. Although not high in magnitude, such a result demonstrates that alcohol use is a risk more tolerated by youth. Additionally, the intercepts in this model were significant for all groups except college/post-graduation students. So, the groups are statistically similar regarding their relational/behavioral features. Once more, as the expectation becomes higher, the magnitude of the intercepts increases.

Tables 15 and 16 presents the results of ordinal logistic regressions for risk data variables. For risk data related to experimentation in Table 15, the only variables significant were those related to alcohol and sex relations. In this model, alcohol experimentation is associated with 1.09 the odds of having a high expectation, while sex relations are related to 20.92% the odds of being less likely to high expectations. Thus, once more alcohol data demonstrate to have a positive influence on students' perspectives.

TABLE 15 – Ordinal Logistic Regression - Risk variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Have you ever smoked cigarettes				
No	1.00			
Yes	0.97	0.90	1.05	0.473
Have you ever had alcohol				
No	1.00			
Yes	1.09	1.03	1.15	<0.01
Have you used drugs				
No	1.00			
Yes	0.98	0.86	1.13	0.823
Have you ever had sex				
No	1.00			
Yes	0.79	0.74	0.84	<0.01
Intercepts				
Elem|High	0.04	0.03	0.04	<0.01
High|HigTec	0.12	0.11	0.12	<0.01
HigTec|Coll	0.17	0.17	0.18	<0.01
Coll|PosGrad	0.38	0.37	0.40	<0.01

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school.

Considering frequency risk data 16, significant alcohol was related to higher odds of having high expectations except for the '20 or more days' category. This high-level drink was associated with 35.22% of the odds of having lower expectations. Still, most of the coefficients for alcohol frequency were not significant. On the contrary, the 'number of sexual partners' variable was significant in almost all categories and demonstrated to be related to lower odds of having high expectations. The highest coefficient was five partners, with 31.22% lower odds of having high expectations.

TABLE 16 – Ordinal Logistic Regression - Risk frequency variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Frequency of cigarette use in the recent 30 days				
Never smoked	1.00			
none	1.07	0.98	1.16	0.120
1 or 2	0.75	0.64	0.87	<0.01
3 to 5	0.71	0.50	1.01	0.057
6 to 9	0.91	0.66	1.27	0.597
10 to 19	1.00	0.65	1.56	0.988
20 or more	0.90	0.64	1.25	0.524
Frequency of alcohol use in the recent 30 days				
Never took	1.00			
none	1.15	1.08	1.22	<0.01
1 or 2	0.95	0.88	1.03	0.200
3 to 5	1.14	1.00	1.30	0.052
6 to 9	1.08	0.91	1.28	0.371
10 to 19	1.20	0.96	1.50	0.110
20 or more	0.65	0.51	0.83	<0.01
Frequency of drug use in the recent 30 days				
Never used	1.00			
none	1.14	0.97	1.33	0.103
1 or 2 days	0.96	0.78	1.17	0.672
3 to 5 days	1.13	0.76	1.69	0.537
6 to 9 days	1.01	0.69	1.48	0.940
10 or more	0.82	0.53	1.27	0.375
Number of sexual partners				
none	1.00			
1	0.87	0.80	0.95	<0.01
2	0.72	0.64	0.80	<0.01
3	0.82	0.72	0.92	<0.01
4	0.86	0.71	1.03	0.100
5	0.69	0.56	0.84	<0.01
6 or more	0.78	0.69	0.89	<0.01
Intercepts				
Elem|High	0.04	0.03	0.04	<0.01
High|HigTec	0.12	0.11	0.12	<0.01
HigTec|Coll	0.17	0.17	0.18	<0.01
Coll|PosGrad	0.39	0.37	0.40	<0.01

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school.

As not knowing the expectation could not be considered in the ordinal logistic model, auxiliary regressions are carried out to verify characteristics associated with being indecisive about schooling prospects. In these models, not having an expectation is labeled as one, and having any expectation is zero. Table 17 presents the binary logit model for socioeconomic variables. The features associated with higher odds of not knowing their expectation are being male (OR=1.13), from public schools (OR=1.24), of black color (OR=1.09). At the same time, having a cellphone (OR=0.88), and internet at home (OR=0.80) decreases the odds of not knowing their expectations. Also, the number of people at home increases the odds of not knowing the expectation by 5.00% for each additional person, holding all other variables constant.

Table 18 returns the binary logit results for relational/behavioral variables. For this set of variables missing classes without parents' permission was associated with higher odds of not knowing their expectations. The other significant variables were mostly related to a decrease in the odds of not knowing the expectation. They were having the HPV vaccine (OR=0.88), parents that understand their problems (OR=0.85), kind colleagues (OR=0.80), suffering bullying (OR=0.87), and having friends that drink alcohol (OR=0.92).

The risk data experimentation model (Table 19) followed the same pattern from ordinal regression, with alcohol and sex variables statistically significant and opposed in meaning. Alcohol experimentation was related to 1.15 times the odds of not knowing expectation while having sexual intercourse was related to a 9% decrease in the odds of not knowing the expectation. The binary logit model for risk frequency (Table 20) had mostly not significant coefficients. For that significance, alcohol frequency was related to higher odds of not knowing the expectation and the number of sexual partners related to higher odds of knowing.

The results in this section indicated some features that can be pointed out to be more related to high expectations. An example, being female, from capital cities, with internet access, among other variables. Nevertheless, testing these results is an important fashion of securing the analysis. The main cause of concern to robust analysis was the adequation of the data to the ordinal logistic regression model. As stated in the identification strategy section, the main assumption is the proportional odds approach for the model applied in this paper. This strategy fits cumulative link models, so the proportional odds assumption must be tested to verify the reliability of the model. If the proportional odds assumption is held, each model's coefficients should be approximately the same. This means that the model's coefficients cannot highly alter once response variables are added because the slopes of the outcomes are assumed to be the same.

To test the proportional odds assumption binary regressions were done adding expectation groups. Starting with elementary expectations, each new regression considered the next expectation until post-graduation. Here, a visual plot is used to analyze if the proportional odds assumption holds all beta coefficients should approximately be the same. Figures 4 to 7 indicate the robust check, with the higher variation being

TABLE 17 – Binary logistic regression for "Do not know- Socioeconomic variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Intercept				
Gender				
Girls	1.00			
Boys	1.13	1.06	1.20	<0.01
City				
Capital	1.00			
Interior	1.02	0.94	1.10	0.616
Region				
North	1.00			
Northeast	1.14	1.05	1.24	<0.01
Southeast	1.18	1.06	1.30	<0.01
South	1.25	1.10	1.41	<0.01
Midwest	1.17	1.06	1.29	<0.01
School Type				
Private	1.00			
Public	1.24	1.13	1.37	<0.01
Skin Color				
White	1.00			
Black	1.09	1.00	1.19	0.043
Yellow	1.12	0.97	1.29	0.139
Brown/mixed-Skin	0.99	0.91	1.06	0.707
Indígenous	1.05	0.90	1.24	0.507
Has a Cellphone				
No	1.00			
Yes	0.88	0.81	0.96	<0.01
Internet				
No	1.00			
Yes	0.80	0.74	0.86	<0.01
Has a Car				
No	1.00			
Yes	0.95	0.88	1.02	0.141
Housekeeper				
No	1.00			
Yes	0.90	0.80	1.02	0.089
Lives with the Mother				
No	1.00			
Yes	0.94	0.86	1.03	0.215
Lives with the Father				
No	1.00			
Yes	0.99	0.93	1.05	0.668
People in the house	1.05	1.03	1.07	<0.01

Source: Elaborated by the author using PeNSE data.

TABLE 18 – Binary logistic regression for "Do not know- Relational/Behavioral variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Intercepts	0.32	0.26	0.39	<0.01
Parents smoke				
No	1.00			
Yes	1.06	0.99	1.14	0.106
Missed classes				
No	1.00			
Yes	1.07	1.00	1.15	0.068
Body Perception				
Very Thin	1.00			
Thin	0.90	0.78	1.03	0.128
Normal	0.93	0.82	1.06	0.291
Fat	0.94	0.80	1.10	0.426
Very Fat	0.93	0.71	1.22	0.618
Feels Lonely				
No	1.00			
Yes	0.97	0.89	1.05	0.430
HPV Vaccination				
No	1.00			
Yes	0.88	0.81	0.95	<0.01
Comprehensive parents				
No	1.00			
Yes	0.85	0.80	0.90	<0.01
Well-Treated by the Colleagues				
No	1.00			
Yes	0.80	0.75	0.85	<0.01
Practiced Vitimization				
No	1.00			
Yes	1.03	0.94	1.12	0.552
Suffered Victimization				
No	1.00			
Yes	0.95	0.88	1.03	0.229
Suffered Bullying				
No	1.00			
Yes	0.87	0.82	0.93	<0.01
Close Friends	0.99	0.96	1.03	0.777
Alcohol Friends	0.92	0.89	0.95	<0.01
Drug Friends	1.02	0.97	1.07	0.40
Victim of sexual violence				
No	1.00			
Yes	0.99	0.86	1.15	0.934

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school.

TABLE 19 – Binary logistic regression for "Do not know- Risk variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Intecepts	0.16	0.15	0.18	<0.01
Have you ever smoked cigarettes				
No	1.00			
Yes	1.15	1.04	1.26	<0.01
Have you ever had alcohol				
No	1.00			
Yes	1.04	0.97	1.12	0.283
Have you used drugs				
No	1.00			
Yes	0.92	0.81	1.05	0.232
Have you ever had sex				
No	1.00			
Yes	0.91	0.85	0.98	0.013

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school.

found in Figure 7, that considers risk frequency variables. Even for risk frequency data, variation was not quite large and the coefficients were approximately the same.

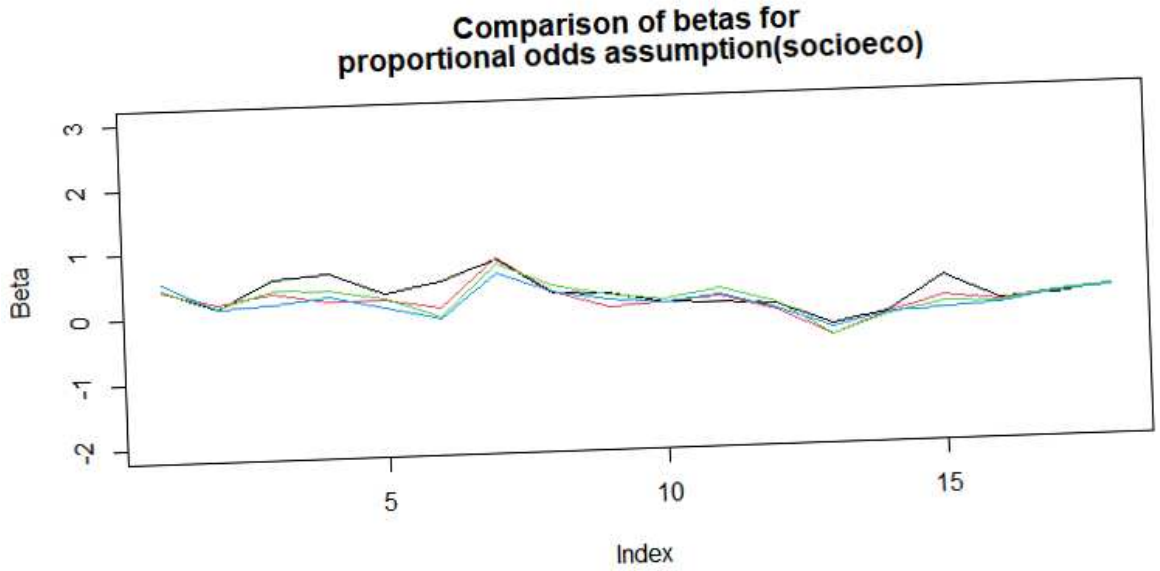

FIGURE 4 – Proportional odds assumption for socieconomic data

Furthermore, F-tests and Chi-square tests were made to analyze individual significance for all models. The results are presented in the Appendix (Tables 55 - 62). For ordinal logistic regressions, living with the father was the only socioeconomic variable not significant (Table 55). Regarding the relational/behavioral model, three variables were not significant. They are ‘suffered victimization’, ‘victim of bullying’, and ‘drug use by friends’ (Table 56). Therefore, even though these variables are significant in the model, their significance does not hold after a robustness check. For risk data

TABLE 20 – Binary logistic regression for "Do not know- Risk frequency variables

Variables	Schooling Expectation			
	OR	IC 95%		p-value
		Inferior	Superior	
Intecepts	0.16	0.15	0.18	<0.01
Frequency of cigarette use in the recent 30 days				
Never smoked	1.00			
none	1.15	1.04	1.26	<0.01
1 or 2	1.05	0.88	1.26	0.576
3 to 5	1.47	1.04	2.06	0.028
6 to 9	1.45	0.90	2.35	0.130
10 to 19	0.72	0.45	1.14	0.156
20 or more	1.29	0.90	1.85	0.169
Frequency of alcohol use in the recent 30 days				
Never took	1.00			
none	1.01	0.94	1.09	0.749
1 or 2	1.14	1.04	1.26	<0.01
3 to 5	0.96	0.79	1.16	0.659
6 to 9	0.96	0.77	1.18	0.681
10 to 19	0.96	0.74	1.25	0.772
20 or more	1.07	0.82	1.41	0.611
Frequency of drug use in the recent 30 days				
Never used	1.00			
none	0.89	0.75	1.05	0.167
1 or 2 days	0.86	0.67	1.10	0.225
3 to 5 days	1.00	0.71	1.40	0.995
6 to 9 days	1.24	0.72	2.12	0.443
10 or more	1.12	0.74	1.69	0.602
Number of sexual partners				
none	1.00			
1	0.90	0.82	0.99	0.026
2	0.85	0.74	0.98	0.022
3	1.06	0.91	1.24	0.445
4	1.09	0.88	1.35	0.448
5	0.85	0.64	1.12	0.242
6 or more	0.82	0.70	0.96	0.013

Source: Elaborated by the author using PeNSE data.
Adjusted for sex and type of school.

variables alcohol experimentation was not significant (Table 57), so the only result that remained reliable for this model was the sex relations variable. Risk frequency data had all variables significant (Table 58).

The binary estimations presented also had not significant variables. The socio-economic model had the capital, car, living with a father, and living with a mother as non-significant (Table 59). The relational/behavioral model for ‘do not know expectations’ presented many non-significant variables. They were parents smoke, body perception, feeling alone, colleagues treat well, practiced victimization, bullying, friends that drink alcohol, and sexual violence (Table 60). Thus, this model is not as reliable as the ordinal method we used. Finally, for risk data (Table 61), cigarette and sex relations variables were significant, while for risk frequency data, alcohol and sex partners were significant

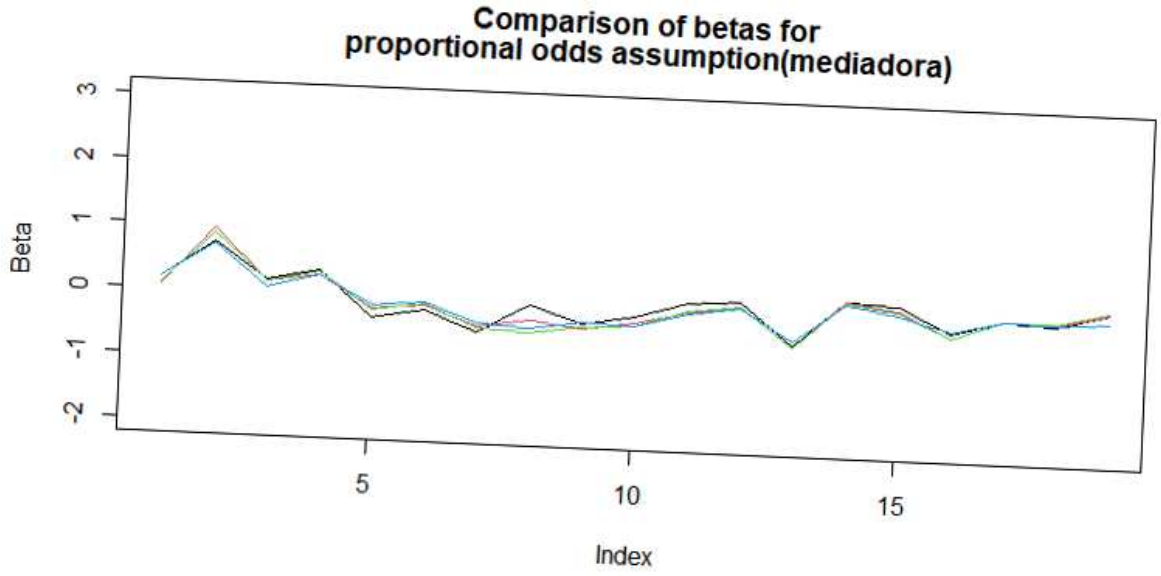

FIGURE 5 – Proportional odds assumption for relational/behavioral

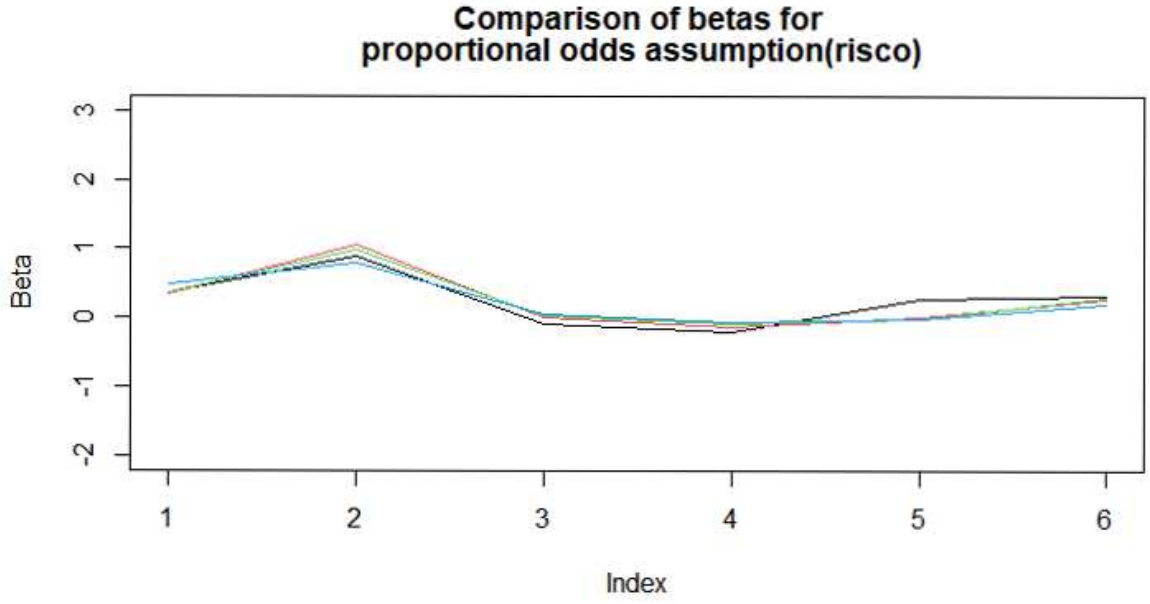

FIGURE 6 – Proportional odds assumption for risk.

(Table 62). The next section discusses the main results observed in this section and relates them to the literature.

1.5 DISCUSSION

The main purpose of this study was to analyze subjective expectation data in association with related variables. Models were drawn for socioeconomic, relational/behavioral, and risk data variables. Our first analysis from the results is that subjective expectation data can be a powerful source of analysis concerning youth and

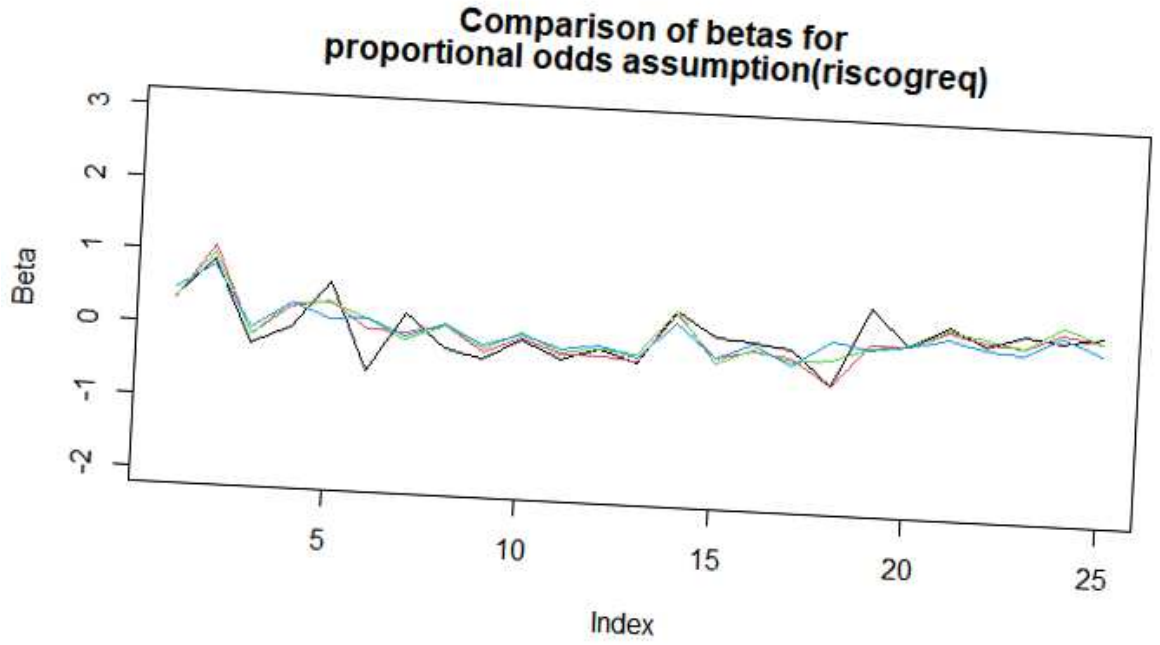

FIGURE 7 – Proportional odds assumption for risk frequency

should be considered in designing policies regarding the youth.

A common result of the models was that female student have a higher expectation compared to males. This result is also seen in the literature and might address that girls anticipate gender discrimination in the labor market, and/or use schooling as a fashion of empowerment (LAUGLO; LIU, 2019). Also, Favara (2017) mentioned, the opportunity costs of staying in school might increase relatively more for boys than for girls. As men usually start to work earlier, girls might be more flexible in combining studying with other household responsibilities usually designated to them. Thus, gender stereotypes might explain why girls are more likely to expect to stay in school than boys.

Capital students were also more prone to have high expectations. As interior regions are not as economically developed as capital cities, perhaps the returns and opportunities from schooling in capital cities are more evident to students, which addresses this feature in their expectations. Also, the type of school demonstrated to be an important feature to students' schooling prospects. In this case, students from public schools had high odds of having lower expectations than those from private schools. This result is likely to express the structural problems elementary education faces in Brazil. Basic education still reflects deficits and is a disadvantage compared to a private school in the Brazilian context (SCHWARTZMAN; BROCK, 2005).

The skin color was related to lower expectations for black and indigenous students. But not for brown/mixed skin students. This result indicates that racial subjects might also impact youth's perceptions, especially when the results most significant were for black students. On the other hand, brown/mixed skin students are composed of a widely mixed population, which could explain the insignificance in this result.

Interesting results were those related to economic condition, having a cellphone, internet at home, and someone in the family having a car increased the odds of having high expectations. These results can be analyzed as proxies to the economic situation and indicated that there might be a strong influence of these subjects on student's prospects. Special attention is drawn to internet access, which can dramatically change students' performance, help with their studies, and it had the highest association.

Moreover, our results demonstrated that living with the mother is associated with higher expectations, but living with the father was not. Such a result aligned with the economic variables analyzed might support Heard (2007) analysis that parent presence matters a great deal, but family stability matters even more. In this sense, instability in the socioeconomic condition and family relations can have negative consequences worse than living with the father's absence. Also, the number of people at home be associated with higher expectations could be related to this variable assimilating large families with a vulnerable economic condition that might need to enter the labor market earlier.

Therefore, the socioeconomic analysis indicated that economic condition is likely to play an important role and family structure largely matters. Once small differences among children in the preparation provided by their families can be frequently multiplied over time into large differences when they are teenagers (BECKER, 1993). So, this is likely to affect youth's expectations. Still, the analysis of relational/behavioral variables pointed out that students with parents that smoke and miss classes without parents' approval have higher odds of having lower expectations. On the contrary, bullying-related variables did not hold the results after a robustness check. (CORNELL et al., 2013) mentioned that teasing and bullying might affect dropout rates and the whole school environment. Nevertheless, our model was not able to capture such dynamics.

Furthermore, having the HPV vaccine, comprehensive parents, and being well treated by their colleagues are associated with higher odds of high expectations. The HPV vaccine is highly significant due to a prevalence of vaccination among females, who have higher expectations than boys. Also, it could indicate that precautioned people are more prone to develop high expectations because they expect more schooling returns. Being well treated by colleagues and parents that understand their problems contributes to an environment where the student feels accepted, and this is likely to contribute to a better prospect from schooling.

Regarding risk data, sex relations variables were more related to a lower schooling expectation. As risk sexual behavior might contribute to unwanted pregnancy and STDs (ALMEIDA; ARAÚJO-JÚNIOR, 2016) that have serious consequences, they can probably be better integrated into students' expectations. Still, not much more can be said once these models' results were not reliable for all variables. In general, the results pointed to a prevalence of higher expectations among female students, from capitals, of white color, studying in private schools with access to economic possessions, such as cellphone and the internet. Additionally, an environment of acceptance, without addiction

problems is likely to contribute to student's expectations.

The cumulation of disadvantage concept could help to understand this context. Designed by Elder (1998), this concept draws attention to the effect adverse environments might have on people's perspective, being cumulative over time. Such factors that revolve in a negative environment might have serious consequences for youth's well-being, reducing their prospect of schooling could be one case. Still, low expectations raise a concern about the student's long-run welfare.

A usual subject that comes to discussion regarding schooling is financial aid. To this matter, (FULLER; MANSKI; WISE, 1982) confirmed that financial aid can be an important determinant of postsecondary school attendance. Financial aid is also applied in Latin countries. The Mexican program Oportunidades, for instance, finances youth studies and raised the possibility to pay part of their schooling grants directly to the youths to improve intervention's effectiveness and school attendance (ATTANASIO; KAUFMANN, 2014).

Also, Chiapa, Prina e Parker (2016) provided evidence that financial access increased the schooling level of daughters and the educational aspirations and expectations parents hold for their children in Nepal. In Brazil, there are policies related to financial aid in college (FIES, PROUNI) and public universities figures as the highest in concept and quality. Still, seems to exist some structural factors that prevent these youth from having a higher expectation. For instance, maybe they do not know about these possibilities, do not see a return from schooling, or are forced to enter the labor market due to income needs.

Moreover, though the normative element is not the concern of this paper, we raise the question of expectations are needed to be high. PeNSE data has some variation, but a large set of students hold high expectations. As (SMITH-GREENAWAY; YEATMAN, 2020) stated,

> ...the rapid expansion of schooling across low-income countries, combined with intensive governmental and nongovernmental efforts to promote education, has encouraged youth in these contexts to form exceptionally high educational expectations, despite immense structural barriers to achieving them. Consequently, many young people's educational expectations go unmet, driving concerns over the possible unintended consequences, including their elevated risk of mental health problems. (Smith-Greenaway; Yetman, pg 1, 2020)

In this context, unmet schooling expectations should also be a source of concern, as negative consequences can be verified. Besides, students might treat expectation as aspiration, with answers not corresponding to what they think will happen but to what they hope will happen. Such bias does not invalidate our analysis but is an important note to remember to avoid misleading interpretations. As a result, expectations might highly shift during adolescence. Due to the endogenous nature, the prospect has, individuals

might update their efforts and perspectives when they learn that the outcomes are inconsistent with their prior expectations (FAVARA, 2017).

In this sense, policy applications should not focus on raising expectations but on creating an environment where high expectations are possible to be accomplished. Work social capital forms, such as information channels, norms, and closure of social networks might play an important role in students' perceptions and educational achievement (COLEMAN, 1988). Family and school-level policies are critical, once parents are the main reference children have during most of their lives and teachers can highly influence students' perceptions.

Zhang (2011) analyzed that teachers' perceptions of the importance of children's family background and economic condition can be strongly associated with teachers' evaluations and expectations, helping to predict children's later school persistence with substantial influence on school outcomes. Consequently, there are benefits from closer connection and better understanding between their teachers and parents.

Further, Weinstein, Madison e Kuklinski (1995) verified that policies aimed to make substantive changes on students' perceptions resulted in negative expectations disconfirmed and new behaviors acknowledged. However, the authors claim that creating a positive expectancy climate requires integrative changes in the curriculum and group instruction, focusing on attitude/stereotype change and school reform efforts to address both psychological and systemic change.

Such analysis could also be applied to students with no expectation, to the extent that might help them formulate what they want for their future. Our results indicated that the features related to not knowing the expectation were similar to those related to lower expectations, so this group could have a propensity to confirm expectations to be lower at some moment in their lives.

Finally, this study has several limitations. First, it is a correlational study, to the extent that it is not possible to make any causal inference. Second, it is a cross-section study, with no elements regarding time being possible to be analyzed. Third, different from other studies, this paper could not consider parents' expectations of youth. Although the adolescent's decisions have a self-perception level, they can be highly influenced by family perspective, with important implications not accounted for in this study. Fourth, we cannot analyze how students' expectations respond to expect income returns once our subjective expectation data broadly questioned youth about which level of education they aim to have. Such aspects require future research.

1.6 CONCLUSION

This study aimed to analyze subjective expectation data associated with socioeconomic, relational/behavioral, and risk variables. Our results indicated that unfavorable socioeconomic situation, relational/behavioral, and risk aspects are significant aspects in association with lower schooling prospects. The central interpretation from this paper

is that practical policies should focus on creating an environment where students' high expectations can be achieved. Also, efforts to improve school structure, and policies to reinforce the closure of family connections are a relevant fashion that might help students form expectations and think about their long-run welfare.

As our risk data results were not quite clear, the next chapter implements an investigation of risk data in association with schooling expectation based on an alternative theoretical and empirical approach.

2 HIGHER RISKS, LOWER EXPECTATIONS? AN ANALYSIS OF THE RELATIONSHIP BETWEEN RISK ENGAGEMENT AND SCHOOLING EXPECTATION

2.1 INTRODUCTION

Choice analysis within economics is frequently based on weighing the benefits and costs. Such analysis could also be applied to risk behavior and may explain a wide range of behaviors that might have negative outcomes on health. When studying risk behaviors among young people, this issue becomes imperative. Youths engage in a host of potentially risky behaviors, which have important implications for their well-being and life prospects (GRUBER, 2001). Those behaviors can include smoking, drinking, and having unprotected sex, and these behaviors are generally seen before age nineteen (O'DONOGHUE; RABIN, 2001). Weighting the potential costs and benefits of such behaviors leads to the reasoning that most of these activities have a low benefit and high costs.

As stated by O'Donoghue e Rabin (2002), two characteristics can be addressed to risk behaviors. First, the more of the good a person has consumed in the past, the lower is his well-being now. Second, the more of the good the person has consumed in the past, the more he desires now. This situation might lead to a trap of addiction, although the person has less pleasure from the consumption it continues to consume because the burden of stopping hurts even more. The implications of risk behavior might appear as determinants of fertility and transmission patterns of sexually transmitted diseases (STDs) (WELLINGS et al., 2006), involvement in criminal activities (DHAMI; MANDEL, 2012), among other factors. Also, it is worth noting that the long-term costs of such activities may include adverse impacts on employability, longevity, marital stability, and other results (NELSON, 2010).

To understand these behaviors, psychological studies indicate that the obstacle would be in the problem of self-control, suggesting that people are rational about their future behavior, but have difficulty using it to adapt current preferences (FISCHER, 1999). A contribution to the theme is models developed in psychology, which suggest several dimensions along with economic models can deepen to understand young people's decision making (GRUBER, 2001). In addition to this, behavioral economics can bring good insights to the subject, once the field seeks to analyze the preferences and decisions of agents through economic models that incorporate psychological aspects of individuals' decision-making process.

In general, it is emphasized that research on risk behaviors generally finds evidence consistent with a perspective of bounded rationality. It relies on the understanding that youth's decision to engage in a risky activity is associated with the perceived benefits, with scarce evidence to suggest that they consider and integrate the costs involved (DHAMI; MANDEL, 2012). Although experts correlate risk analysis with techni-

cal estimates of annual fatality, laypeople rely on other hazard characteristics and can highly deviate from experts' perspective (SLOVIC, 2000b). Public perceptions derive from intuitive associations, economic interests and reflect cultural values (KASPERSON et al., 2000). Accordingly, when it comes to youth, concerns with identity formation and peer pressure effects play an important role (STEINBERG, 2004).

This paper aims to analyze risk behavior based on models developed in behavioral economics. The multiple-motive approach is addressed, allowing the interpretation of risk behavior considering the effect of anticipatory utility, visceral factors, hyperbolic discounting, and other concepts. We analyze the risk behavior related to schooling expectations of young people, considering the school prospect as a reference on how young people aspire to their quality of life in their future. Analytical instruments based on behavioral economic models support the understanding of students' decision processes facing risky activities.

Furthermore, the empirical methodology is based on structural equation models, used to obtain latent constructs that measure the propensity to engage in different types of risk. Aspects of sexual, alcoholic, and substance use are incorporated, and several mediator effects are tested. The data used is the 2015 National School-Based Health Survey (PeNSE), conducted by the Brazilian Institute of Geography and Statistics (IBGE) based on the parameters suggested by the World Health Organization (WHO). The variables used for the formation of risk constructs consider the recent consumption of alcohol, nicotine, marijuana, and crack, as well as variables referred to the age of initiation of alcohol use, nicotine, drugs, first sexual activity, sex practice without a condom and number of sexual partners (IBGE, 2016).

The article contributes to the literature in three aspects. Firstly, by analyzing the relationship between risky behavior, schooling expectations, and mediation effects for a developing country with high cultural and regional diversity. Second, by allowing behavioral inferences from an extensive database. And finally, by exploring behavioral factors related to heuristics and biases that may indicate the path of public policies to discourage risky behavior among young people. The article consists of four sections, in addition to this, the first section reviews the literature, divided into theoretical and empirical. Soon after, the second section describes the methodology, data analysis, and identification strategy. The results obtained and their external validity is discussed in the third section, and the fourth section concludes.

2.2 LITERATURE REVIEW

Human capital investment is the main fashion to integrate the market with higher earnings. Theorized by Becker (1993), the human capital analysis assumes that schooling raises productivity and earnings by providing knowledge, skills, and ways of analyzing problems. Besides, the concept of human capital can aggregate expenditures and efforts on medical care. It happens because it is not possible to separate a person

from her knowledge, skills, health, and values. As a result, these features may provide a raise in human capital stock (BECKER, 1993).

Relative to the youth investment in human capital, a serious basement to young achievement is the family structure. Parents have a large influence on education, marital stability, and other dimensions, in which families with low education, welfare dependence, early pregnancy, and marital instability are described as “underclass” (BECKER, 1993). Beyond that, these vulnerabilities might lead to a young student dropping out of school, which is a severe problem for young people’s investment in human capital. Understanding the process of expectation might be an important path to understand students’ triggers to a lower expectation. This consists of an important measure once the literature supports that young people who drop out of school are unlikely to ever return (CARD; LEMIEUX, 2001).

The first paper aimed to analyze the main characteristics regarding schooling expectation. Still, it was not clear how risk behaviors might be related to schooling expectations. Considering the engagement in those activities as a fashion to reduce the human capital stock, this article aims to understand why young people involve in such behaviors and analyze it in light of the economic literature regarding risk behavior and schooling expectation. We consider that the subjective data of schooling expectations is a good manner of representing people’s prospects and might be related to risk behavior in different ways.

Problems generated by risky behaviors may generate economic consequences, as productivity losses can occur due to the health problems that, in turn, influence individuals’ outcomes, their perspectives, and quality of life (HEINECK; SCHWARZE, 2003). Thus, studying risky behaviors in early adolescence, when such activities tend to appear, has important public policy implications, with potential long-term effects.

In this context, behavioral economics provides new conceptual systems to inform the scientific understanding of behaviors in different fields, such as health for example. Moreover, it seeks to translate scientific understanding into practical and effective behavior change (BICKEL; MOODY; HIGGINS, 2016). The focus of the present study is on the application of economic behavioral models based on theoretical instruments. Nevertheless, rational choice analysis is also considered, assisting the study of risk behavior among young Brazilians. Human capital analysis and individual risk behavior are approached in theoretical terms using inconsistencies and biases in the following sections. Then, empirical evidence is presented to display the relations between risk engagement and schooling expectations.

2.2.1 Theoretical Background

As risk behavior activities might lead to harmful addictions, we begin by defining addiction. The review implemented by Herrnstein e Prelec (1992) addresses four interpretations of addictive behavior. First, it is a disease, this approach focuses on the

consequences of addiction and not on the decision process that leads to it. Second, the rational approach, with one of the most elaborated economic models presented by Becker e Murphy (1988), considering addiction as an optimal intertemporal plan. Third, the primrose path, with addiction being viewed as a trap in which people are lured because of hidden latent costs. Lastly, the divided-self approach, in which people can make choices that their future selves would not appreciate, but they only notice it as time passes, and the future self comes to play.

The main analysis considered in this paper is the divided self. To understand the economic model regarding decision making with short-term benefits and long-term costs we begin by analyzing time discounting and intertemporal choice. Time discounting encompasses any reason for caring less about a future consequence, this includes factors that diminish the expected utility generated by a future consequence, such as uncertainty (FREDERICK; LOEWENSTEIN; O'DONOGHUE, 2004). The discounting might be explained by an 'excessive myopia', to what people do not see clearly long-term costs and discount future too much, as well as to time inconsistency, in which preferences may vary over time (GRUBER, 2001).

The time-inconsistency subject was first approached in economic models by the seminal paper of Strotz (1955). He demonstrated that any discount beyond exponentially discount would imply time inconsistency, with people being present-oriented. This situation can also be described with the salience concept, characterizing short-term discount rates being higher than the long ones, if salience matters, by making preferences dynamically inconsistent, the implications for economic models are profound (FISCHER, 1999). So far, the most used type of discount with time inconsistency assumption has been the hyperbolic discount, which can address self-control problems that arise from this short-term propensity to pursue immediate gratification that is inconsistent with long-run preferences (RABIN, 1998).

Functions underlying hyperbolic discount have high discount rates for small delays and low discount rates for long delays, i.e., a person may prefer a later reward from a distance, but when the time is near, the more imminent reward might be easily chosen (FISCHER, 1999). This means that a person might consume more than he/she would like a priori. The time-inconsistency discussion matters to the risk behavior because the costs are usually not encountered when youth engage in risky activities. An explanation in terms of cost-benefit is that although risk and benefit may be positively correlated in the environment, they tend to be negatively correlated in the mind of people in this situation. If an activity is seen as 'liked', people tend to judge its risks as low and benefits as high, in a process of affection for the risk, although they might eventually regret the decision (FINUCANE et al., 2000).

Attaining in the second and fourth approaches, the model implemented by Becker e Murphy (1988) relies on a weak concept of rationality, allowing myopic and present-oriented bias as time preference for the present grows. In this model, a strong addiction requires the effect of past consumption on current consumption, so the possibility of

becoming addicted depends on the initial stock of consumption and the demand curve. However, to deal with the possibility of zero initial stock of consumption, the authors consider that some events affect directly the stock, hooking rational people to addictive goods.

This paper seeks to develop an analysis of the relationships between engaging in risky activities and schooling expectations considering students[1] of the ninth grade of Brazilian schools. In such a case, some drawbacks must be pointed out to Becker & Murphy's model. First, this framework brings nothing special about youths relative to adults, although they may behave differently, seek novelty and stimulation, and be more sensitive to prices (GRUBER, 2001). Second, the model provides no formal analysis of why the person would choose to develop this harmful addiction in the first place (O'DONOGHUE; RABIN, 2001). Third, as the model consider the person being rational, do not encounter the slow development of self-regulatory capabilities during adolescence, like impulse control for example (STEINBERG, 2004). As a result, the model works regarding market demand and price fluctuations but has little to explain concerning the start of harmful behaviors on young people.

As an alternative, the youthful model developed by O'Donoghue e Rabin (2001) is a simplified version of Becker & Murphy's model assuming consumption as a binary choice, instead of continuous. This model enables the investigation of the causes young people might indulge despite long-term costs. The basic lesson underlying the model is that the self-control problem is one source of overconsumption of addictive goods, and the awareness of these self-control problems might mitigate or exacerbate this overconsumption.

Some decisions may be recognized as negative to people's long self-interest at a cognitive level, even in the moment of succumbing to the impulse of consumption (LOEWENSTEIN; PRELEC, 1992). Nevertheless, youth can easily be driven by visceral factors, which are situations with a direct hedonic impact that influences the relative desirability of different goods and actions (LOEWENSTEIN, 2004). This occurs because it is easier to say no to a hypothetical act of unprotected sex or a hypothetical beer than to a real moment of pleasure and passion, as well as to a frosty beer right in front of you (STEINBERG, 2004).

The role of visceral factors integrates a wide range of affective influences that are not usually seen at a cognitive level. The study of Loewenstein (2004), who developed this concept, indicates that visceral factors are predictably correlated with external circumstances, such as stimulation and deprivation, however, does not imply a permanent change in a person's behavioral dispositions. He establishes propositions to cover a set of situations that might be under the influence of visceral factors. He establishes propositions to cover a set of situations that might be under the influence of visceral factors. Two of these propositions are directly related to the object of this study.

[1] Mean age 14 years-old.

The first proposition states that the discrepancy between the actual and desired value placed on a particular good or activity increases with the intensity of the immediate visceral factor. The second proposition explains that future visceral factor produces little discrepancy between the value we plan to place on goods in the future and the value we view as desirable. Both propositions imply that people will give greater weight to immediately experienced visceral factors than to delayed visceral factors. In this sense, impulsive behavior might occur when people are driven by high emotional situations that encounter visceral factors, such as hunger, thirst, sexual desire, among others. (LOEWENSTEIN, 2004).

The concept of visceral factors is included in the analysis because the hyperbolic discount, by itself, cannot encounter cases of impulsive choice due to affective feelings and high utility variations. Besides, as stated by Frederick, Loewenstein e O'Donoghue (2004), adding the concept of visceral factors to the hyperbolic discount analysis can account more for the phenomenology of impulsive choices, explaining a wide range of these decisions. Then, hyperbolic discount concerns the relation of immediate rewards and long-term costs, and visceral influences act as a transient effect on immediate utilities. This considered, the next subsection presents the youthful model developed by O'Donoghue & Rabin (2000,2001) additionally to the analysis of visceral factors implemented by Loewenstein (2004).

2.2.2 The model

The divided-self strategy to model time-inconsistent preferences considers a person at each point in time being modeled as a separate agent choosing her current behavior to maximize current preferences, with her future selves controlling her future behavior (O'DONOGHUE; RABIN, 1999). At this point, the extent to what people believe about their future preferences matters the most in this model[2].

Youth might underestimate the future harm of current behavior because they do not recognize the extent of day-to-day fluctuations in tastes and the power of peer pressure when it comes to influencing their decision. Moreover, behaving in a risky way affects the marginal risk accrued from future misbehaving, so the intertemporal context plays an important role (O'DONOGHUE; RABIN, 2001). Although the decision process is the main focus of this economic model, such analysis underlies that the decision is easily influenced by aspects such as identity formation, searching for autonomy, and establishing independence.

Consisting of a simplified version of Becker (1993) applying binary choice, the youthful model becomes more tractable, with the possibility to analyze specific characteristics that trigger risk behaviors. A young person's well-being may be determined by the weighted sum of the instantaneous utility u in the current period t to the future

[2] This section do not aim to present the whole formalization of the youthful model but provide an overview of the theory to support the analysis of risk behavior among young Brazilians. For more information, see O'Donoghue e Rabin (2001)

periods T, represented by a sequence τ. The instantaneous utility of the person $u_\tau(r,.)$ is a function of a set of states S_t that might occur under probability p_τ, with $r \in S$ being a state that occurs a risky activity and $. \in S$ representing all other possible states. Thus, well-being W^t is formally defined as measured by the individual's expected utility of:

$$W^t \equiv \sum_{\tau=t}^{T} \left[\sum_{r \in S} p_\tau(r,.) u_\tau(r,.) \right] \tag{2.1}$$

As visceral factors encounter other forms of visceral response beyond short time delays described by the hyperbolic discounting, such as physical closeness and sensory contact, a term is included in the equation 2.2 to consider a visceral situation. If an individual is driven by a visceral influence when dealing with a state of risky behavior r, there may be a positive variation in the weight of that state, to increase the attractiveness of the instantaneous utility related to the risky activity. The ϕ parameter is the weight, when $\phi > 1$ the individual is in a visceral state, and $\phi = 1$ for all other states. In this context, a discount factor must be applied to the equation:

$$W^t \equiv \sum_{\tau=t}^{T} \delta^{\tau-t} \phi p_\tau u_\tau \tag{2.2}$$

In equation 2.2 δ represents the discount rate applied to all individual decisions (the time-consistent discount). Once multiplied by $\phi > 1$ represents the effect of visceral states. Thus, the rate δ is intensified for any period t. This demonstrates that an individual can always make decisions with a wrong discount rate when influenced by visceral states, but it does not capture the existence of variations in the discount rate at different points in time, that is, time-inconsistent preferences. Thus, equation 2.3 incorporates a parameter β that approximates a hyperbolic discount.

$$W^t \equiv u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} \phi p_\tau u_\tau \tag{2.3}$$

In equation 2.3, the parameter β incorporates excessive myopia alongside the present bias, with β<1. Such biases imply a difficulty to analyze long-term costs and result in individuals more concerned with immediate utility decisions u_t than with future decisions. An additional issue is how the present bias may influence risky behavior, as it depends on the beliefs that young people have about their behavior. In face of risk behaviors, the desirability to do the activity now is greater than their current desire to do in the future. Consequently, a person is more likely to engage in an indulgent activity at the moment of action than would have preferred at a prior moment (O'DONOGHUE; RABIN, 2001).

Therefore, a set of distinctions is made between naive and sophisticated people. A naive person is passive concerning its present bias and visceral states, with $\hat{\beta} = 1$. On

the other hand, a sophisticated person is active regarding its present bias and visceral states, trying to estimate how such characteristics can influence his behavior, so that $\hat{\beta} = \beta$. Finally, this represents the extremes, but people may be partially naive when being active concerning the present bias and visceral states, but underestimating their effects on behavior, with $\hat{\beta} < 1$ and $\hat{\beta} > \beta$ (O'DONOGHUE; RABIN, 2001).

The implications of sophistication and naivete are mainly in response to their biases. Sophisticated people foresee their self-control problems, while naive people structure their preferences as a person with time-consistent preferences. So naive people constantly fail with their prior choices due to the propensity for immediate gratification. Also, how sophisticated people deal with self-control problems can be understood as a game against their future selves, that is, their reaction might configure a strategic option to avoid future selves' bad behavior that they do not have control (O'DONOGHUE; RABIN, 1999; O'DONOGHUE; RABIN, 2001).

It might appear awkward to describe a person analysis as different people in different times. Therefore, it is important to establish that it works metaphorically, to say that our decision can quite vary during different periods. Usually, risk behaviors have immediate rewards and delayed costs, and such aspects can affect sophisticated people. Because sophistication implies a pessimism concerning their future behavior there is a dependence on whether this pessimism increases or decreases the marginal cost of current indulgence. If pessimism decreases the marginal cost of current indulgence sophisticated people might involve in risk behavior due to feelings of inevitability. On the other hand, if pessimism increases the marginal cost of current indulgence sophisticated people might refrain from the involvement even before the start of consumption, which is labeled as an incentive effect (O'DONOGHUE; RABIN, 2001; O'DONOGHUE; RABIN, 2002).

The literature points that the incentive effect mitigates the pessimism effect. Then, sophistication helps the person when knowing about how future misbehavior increases the perceived cost of current misbehavior, encouraging them to behave now. Still, sophistication hurts when knowing about future misbehavior decreases the perceived cost of current misbehavior (O'DONOGHUE; RABIN, 2002). At the same time, naive people are hurt because they are optimistic about the future, which leads them to perceive the small costs of current misbehavior. Also, because they fail at self-management, doing things in the present leads them to do more indulgent things in the future (O'DONOGHUE; RABIN, 2001).

Such analysis has implications not just for health but also for schooling decisions. As an adult, the person probably will care about having a good job with high earnings, but as a youth, all jobs might seem unappealing, so youth may be directed to those jobs with lower earnings. This underestimation may raise the odds of dropping out of school (O'DONOGHUE; RABIN, 2001; GRUBER, 2001). As schooling and health are equally important to human capital investment and considering the empirical evidence relating risk behaviors to school dropout, as França e Frio (2018) and Goulet et al. (2020).

We propose an analysis relating risk behavior to schooling expectations. In this sense, assuming that risky behaviors have negative effects on the individual's future and can affect their education, given the risk activities that individuals have been exposed to until now, their beliefs about their behavior may shape their schooling expectations. Based on the relationships developed, individuals can be classified as sophisticated or naive, as shown in Board 3.

Board 3 – Preference analysis of the youth

		Schooling Expectation	
		Low	High
Risk Behavior	Yes	Sophisticated $\hat{\beta} = \beta$	Naive or partially naive $\hat{\beta} = 1$ ou $\hat{\beta} < 1$ e $\hat{\beta} > \beta$
	No	Naive or partially naive $\hat{\beta} = 1$ ou $\hat{\beta} < 1$ e $\hat{\beta} > \beta$	Sophisticated $\hat{\beta} = \beta$

In short, individuals who incorporate their risk behaviors into their future expectations are considered sophisticated, as they can integrate the future costs of decisions taken in the present. In turn, those whose risk behaviors are not aggregate in future expectations are classified as naive or partially naive. Finally, this analysis investigates the possibility of risk consequences be seen before they occur.

2.2.3 Empirical Evidence

Empirical studies involving the relation between risk behaviors and schooling expectations/achievements are scarce, but some interesting results can be seen in the literature. Sutherland e Shepherd (2001) applied logistic regression to a sample of 4,516 students from English schools, verifying that from those who planned to go on to university only 55.6% reported using any substances compared with 73.5% who did not plan to carry on to university. Besides, the author registered low academic achievement as a risk factor that increases with age for the use of cigarettes and illicit drugs, but not with alcohol, which can be related to the normative use of alcohol in society.

For studies using PeNSE 2016, Almeida e Araújo-Júnior (2016), with a propensity score matching strategy identified that risk factors are related to delay in school progression, especially to students with lower socioeconomic levels. Moreover, Escobar et al. (2020) stated that mental health disorders can lead to dropping out of school and are related to risk behaviors.

School-related factors can be predictive of risky behavior, as pointed out by Cowan (2011), the cost of education can be a relevant factor and can be correlated with behavior through its effect on the perspectives of adolescents' colleges.Mihalec-Adkins

e Cooley (2020) in application to the United States, identified that school engagement is associated with lower magnitudes in externalizing and internalizing problems, providing better self-esteem, and developing social skills.

Some studies take place on the application of experiments to analyze inconsistencies and biases of behavioral economics. An example is Arnett (1994), who constructed a sensation scale and applied it to 116 young people attending high school in Atlanta. The author verified that adolescents reported higher levels of sensation-seeking than adults, and males exhibited more sensation-seeking compared to women. Isaksson et al. (2020), based on structural equation modeling (n=1785) draws attention to the importance of factors such as low self-control, disinhibition, and reward-seeking behavior in alcohol consumption, stating that the early initiation of alcohol use is related to a more comprehensive pattern of externalizing behavior and those delinquent peers should be regarded as risk factors.

Concerning the analysis of perceived risk, analyzing a sample of 76 people Fischhoff, Slovic e Lichtenstein (1979) identified that respondents felt that society should tolerate higher levels of risk for voluntary activities than for involuntary ones. Based on affective risk analysis, Benthin et al. (2000) identified that risk behavior activities, such as smoking cigarettes, marijuana, and sexual intercourse were associated with positive affect, as feeling good and being happy in a sample of 411 students. Further, in this study alcohol use was largely associated with social facilitation. Additionally, Finucane et al. (2000) experiment with 54 people demonstrates that changing people's perception of one attribute (increasing risk) tended to influence ratings on the other attribute (decreasing risk). The inverse relation was explained by people consulting their overall affective evaluation of what they judge as risk and benefit.

Considering the use of cigarettes, Harrison et al. (2020), in an analysis of mental health and substance use of 423 young primary offenders in the northeastern United States, verified through logistic regression that recent smokers exhibit greater behavioral and emotional difficulties, so that smoking may be a strategy to reduce stress. Again, compared with non-smokers, recent smokers had over five times the odds of ever using alcohol and another drug, which calls attention to how risk behaviors are highly correlated between them.

Related to risky sexual behavior, França e Frio (2018), using data from PeNSE (2015) with survival analysis method, pointed out that risky sexual behavior is related to economic problems, such as pregnancy in youth, school dropout, and loss of productivity. The authors also identified that lectures in schools worked for girls to delay sexual initiation, but not for men. A great concern for sexual risk behavior is the use of condoms, Do et al. (2020) in application to five provinces of Vietnam using logistic regression, identified that a higher number of sexual partners is associated with diminished intention to use condoms and higher intention to use of drugs. Besides, in the same study, it is also pointed out that although a significant part showed intention to use a condom in the next sexual intercourse, less than half had used it in the last sexual intercourse, which

demonstrates a strong gap between intention and practice.

Risk behaviors may also be associated with the influence of peers and family members. One argument found is that young people start experimenting risks to facilitate relationships between peers and contribute to autonomy development (ZAPPE; ALVES; AGLIO, 2018). O'Brien et al. (2011) examined the effect of peer presence on reward sensitivity for a sample of 100 students and identified that the presence of peers transformed the choices of late adolescents of 18-20 years old into ones similar to younger and less mature teenagers of 14-15 years old. However, the peer effect can also work with a prosocial role. Walters (2020) in a study with 2,905 youths identified an ameliorative effect of prosocial peers on property crime and drug use. Though it is worth mentioning that the effect was not on the neutralization of bad peer influence, but in the direct effect serving as a risk factor for people with few prosocial peers and as a promotive factor for youth with many prosocial peers.

In the case of risk behaviors among minorities, the literature points out that higher levels of discrimination are associated with smoking and marijuana initiation among young Hispanics in the United States (ROGERS et al., 2020). Additionally, non-sexual physical and emotional abuse and physical and emotional neglect have a positive effect on sexual risk behaviors, depression, and anti-socialization, of a magnitude close to physical and emotional sexual abuse (DIAZ et al., 2020)

Another type of analysis, proposed by Yoon (2020), identifies variations in risk behaviors from different degrees of sociability and popularity within and among groups in schools, verifying that experiences of emotional abuse are key predictors for involvement with problematic pairs, turning them more likely to be characterized as severely antisocial group members.

It should be noted that these applications are mostly international experiences with samples that usually are not quite representative of the population in the study. In Brazil, where social inequalities are important markers of risky behavior, understanding these behaviors can delimit important results on how young people incorporate risks in their expectations. The school environment is also supportive for students and has important implications for the well-being and promotion of the sexual and reproductive health of these young people (OLIVEIRA-CAMPOS et al., 2014). In this sense, this study may provide important insights that may be used for the formulation of public policies that reduce risk behavior, as well as improve school youth perspectives.

2.3 METHODOLOGY

This section presents the methodology used to analyze risk data related to schooling expectations. Here, we move forward from the identification strategy implemented in chapter 1 and we use a method more suitable for risk behavior. The section is divided into two. The first subsection presents the data approach to this analysis and the second establish the strategy carried out to the research.

2.3.1 Data

The data implemented in this study has the same structure as Chapter 1, but some differences came out to the implementation concerning structural equation modeling strategy, with risk variables labeled as constructs. Analyzing the variables used in the model, as was done in Chapter 1, is important to understand the composition of risk constructs. A construct represents a set of variables about a specific aspect that captures latent variables, that is, an aspect that cannot be directly observed, i.e, the risk behavior in our case. The PeNSE data has a set of variables concerning the engagement in risk behavior; therefore, by applying factor analysis, this data could be aggregated into a measurement that involves all the variables into a factor in each risk. The reflective constructs[3] developed in the measurement model are divided into three risk groups: alcoholic behavior, sexual behavior, and substance use.

The alcohol risky behavior construct considers the frequency and intensity of alcohol consumption in the 30 days before the survey. Besides, the student was questioned the number of times he got drunk, if he had any alcohol problems, and his age when started drinking. The risky sexual behavior construct incorporates four variables that encounter the age of sexual onset, the number of sexual partners, and whether he/she used a condom during the last and first sexual intercourse. Finally, the construct related to substance use aggregates variables of consumption of marijuana, crack, and any other substance in the recent 30 days. Besides, it considers the age the student started to use the illegal substance and cigarette frequency as well.

It is worth mentioning that the variables related to nicotine use were not strong enough to compose one single factor. Consequently, it was considered the result from the exploratory factor analysis, as well as how the literature considers this variable related to the other risk constructs. Then, we implemented the cigarette frequency variable in the drug and alcohol construct for the main model, with differences in the implementation of multi-group analysis, which will be better discussed in the next sections. In all constructs, some of the questions bring temporality of 30 days or 12 months, avoiding memory bias[4]. The dependent variable is the subjective schooling expectation given by the question “What is the highest level of education that you intend to conclude?".

In addition to the variables mentioned, it was also used variables of sex, area, type of school, region, and color for the implementation of multi-group models and verification of how risky behavior is related to an expectation of schooling on different strata of the population. Alongside this analysis, a construct of bullying and peer effect is made to test mediation effects. The first one consists of four variables regarding the practice of bullying, victimization, and how the student is treated by colleagues. The

[3] Reflective construct causes changes in the observed variables, while formative construct receives changes from observed variables.

[4] Type of cognitive bias that makes it difficult to accurately recall past events as they move away (FISCHHOFF; SLOVIC; LICHTENSTEIN, 1977)

second consists of two variables describing the number of friends involved in alcohol and substance risk activities. For the mediation analysis, variables include age, number of people in the household, internet, car, housekeeper, if parents smoke, and if the student is a bullying victimizer. Board 4 lists all the variables used in the risk factor analysis.

Board 4 – Variables applied in the factor analysis

	Variable	Abreviation	Type	Description
Groups	Sex	sex	Categorical	Male or Female.
	Area	area	Categorical	Capital or Interior.
	Type of school	school	Categorical	Public or Private.
	Region	reg	Categorical	North, Northeast, Midwest, Southeast, South.
	Color or Race	color	Categorical	Yellow, White, Indigenous, Brown/mixed, Black.
Response	Schooling Expectation	expec	Numeric	Scale from 1 to 5 for the expectation, varying from elementary school to post-graduation.
Alcohol	Age Alcohol	agealc	Numeric	Scale from 1 to 7, ranging from the beginning of alcohol consumption at 7 years of age to 18 or over, 1 for those who never drank.
	Alcohol Frequency	alc_freq30	Numeric	Scale from 0 to 7 for days that consumed at least one glass of alcohol, ranging from no glass of alcohol in the last 30 days to consumption of at least one glass in every 30 days.
	Alcohol Intensity	alc_int30	Numeric	Scale from 0 to 7 for intensity of consumption on the days you drank, ranging from no glass of alcohol in the last 30 days to consumption of 5 or more glasses per consumption day in the last 30 days.
	Drunk	drunk	Numeric	Scale from 0 to 5, ranging from zero to 10 or more times in life.
Sexual relations	Age of first sexual intercourse	age_sex	Numeric	Scale from 1 to 6 for the initiation of sexual activity, varying from 9 years old to 18 or more. 1 for those who never practiced.
	Sex Partners	sexpartn	Numeric	Scale from 0 to 6, ranging from no partner to 6 or more.
	Preservative	preserv	Binary	1 = Sexual intercourse without a condom, 0 = With a condom or no sexual intercourse
Drugs	Illegal Drug Age	drug_age	Numeric	Scale from 1 to 7, ranging to the beginning of consumption. Representing 7 years up to 18 or more, 1 for those who have never used it.
	Cigarette Frequency	cig_freq30	Numeric	Scale from 0 to 7 for days you have consumed cigarettes in the last 30 days, ranging from no cigarettes in the last 30 days to consumption in all 30 days. 0 = Never smoked.
	Marijuana Frequency	mar_freq30	Numeric	Scale from 0 to 4 for days that you have used marijuana in the last 30 days, ranging from no use in the last 30 days days until consumption in 10 days or more. 0 = Never smoked.
	Crack frequency	crac_30	Numeric	Scale from 0 to 4 for days that consumed crack in the last 30 days, ranging from no use in the last 30 days to consumption in 10 days or more. 0 = Never smoked.

Source: Elaborated by the author using PeNSE data.

2.3.2 Identification Strategy

Structural Equation Modeling (SEM) is a family of statistical models that seek to explain the relationships between a varied set of variables to examine interrelations expressed in multiple equations (HAIR et al., 2009). Thus, SEMs typically include two or more equations in the model, differing from the usual single equation regression model that has a single dependent variable and multiple covariates (BOLLEN; NOBLE, 2011). As a result, the main purpose of SEM is to describe relationships between constructs, providing a powerful instrument to deal with highly correlated variables.

The constructs in SEM reveal dimensions previously not observed by the manifest variables accurately measuring key concepts. Using this model, interdependence and dependency analysis can be implemented altogether, considering that is possible to

apply confirmatory factor analysis and multiple regression in the same model (HAIR et al., 2017). Accordingly, SEM can quantify and test hypotheses on relations among latent and observed variables, also providing tests of consistency and plausibility and enabling a direct analysis, as well as mediated relationships (BOLLEN; NOBLE, 2011).

The main rationale for building the model of structural equations lies in the development of a theory that explains the relationships between the reflexive constructs. Thus, the model is divided into two parts, one is for measurement and the other structural. In the measurement model, reflexive constructs are created from the manifest variables. In the structural model, the directions of the constructs are traced to verify the theory and analyze the regression. Considering the theoretical structure developed in the study, the directions range from risk behaviors to the expectation of schooling, aiming to understand the relationships between risk engagement and school prospects.

The analysis is commonly represented visually, in Figure 8 as the measurement model and the structural model is established. The measurement model[5] reflects the formation constructs, represented by circles pointing to the manifest variables, represented by squares. The structural model reflects the relationships of the constructs, with the arrows pointing from risk constructs to the schooling expectation. As PeNSE data has risk variables highly correlated (Figure 3, Chapter 1) the measurement model of SEM is quite useful. Thus, variables related to the same risk are combined to create a factor that encounters the features of the risk of alcohol, drug, and sex relations. Figure 8 plots the model.

To verify if a dataset is adequate to apply factor analysis, a common indicator used is Kaiser-Meyer-Olkin (KMO) factor adequacy. KMO tests can test each variable as well as the whole dataset of the group and represents a measure of communality among variables that might divide the same proportion of variance (VOGT, 2005). To ensure a better acceptance of factor analysis, the values must be above 0.70. Table 21 presents the test, KMO is above 0.80 for all variables and the overall factor adequacy relies on 0.89, which indicates that risk data is an adequate set to factor analysis implementation.

TABLE 21 – Kaiser-Meyer-Olkin factor adequacy

Adequacy for each item / Overall Adequacy = 0.89						
idadealc	alc_freq30	alc_int30	bebado	probeb	cigar30	idadedrog
0.88	0.88	0.88	0.94	0.9	0.98	0.88
droga30	maconha30	crack30	idadersex	preserv	preserv1	pessrsex
0.84	0.83	0.89	0.93	0.84	0.84	0.96

Source: Elaborated by the author using PeNSE data.

Another concern regarding method application is the technique used to fit the model. A few exploratory factor analyses[6] (EFA) were carried out to analyze the response of the data to different techniques. Ordinary least squares (OLS), weighted least squares

[5] We used R package for latent variable analysis developed by Rosseel (2012), for the complex survey we used Oberski (2014) and Epskamp (2019) for plotting.

[6] The 'psych' package was used (REVELLE, 2019).

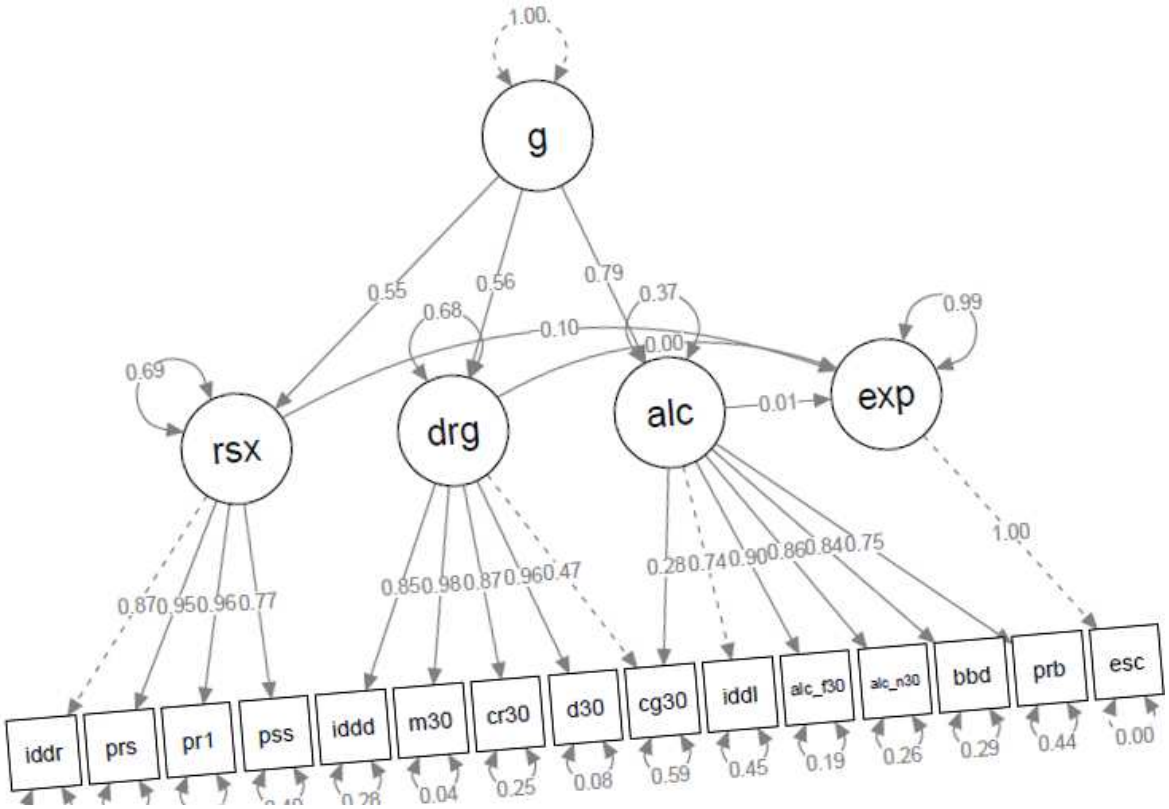

FIGURE 8 – SEM Path

(WLS), generalized weighted least squares (GLS), and maximum likelihood (ML) were tested to verify which was more acceptable. Of those, maximum likelihood had better indices, with a reduction on Chi-square and BIC[7] criteria (Appendix 63). Therefore, the ML estimator was the technique applied in the study.

Similarly, ML choice is aligned with SEM literature, which usually has this estimator as preferred, due to the well-behaved matrices found when applying this technique (REVELLE, 2019). As maximum likelihood estimation is a good technique for estimating parameters and drawing statistical inferences in a variety of situations, especially nonstandard ones, it is reasonable to consider this method also based on the theory (HARRELL, 2015). Equally, different rotations were applied in ML estimations of EFA tests (Tables 64 - 67 from Appendix), as well as hierarchical scheme (Tables 68 - 71, Appendix). Such analyses supported the choice of the variables to the formation of constructs, giving statistical reliability, which was aligned with the theory precluded. Thus, the base model applied of the paper is,

$$
\begin{aligned}
& semmodel = \\
sexr = \ & agesex + preserv + preserv1 + sexpartn \\
drug = \ & cigar30 + agedrug + marj30 + crack30 + drug30 \\
alc = \ & agealc + alcfreq30 + alcint30 + drunk + probalc \\
expec = \ & schoexp
\end{aligned}
$$

[7] Lower values mean better performance to the models, good to analyze in comparison to other models, once there is no reference number for this test. The same applies to Akaike's information criteria (AIC).

$$expec = \ alc + sexr + drug$$

As we applied weights to the sample, we tested this model to verify the difference in the fit indices (Table 70). The main effects were seen in the BIC, AIC, and RMSR[8] indices, in which a reduction was verified. Regarding the cigarette frequency variable, though the statistical analysis claimed to be more adequate with drug construct, the theory occasionally associated its dynamic with alcohol. In such context, we tested the cigarette variable being in the alcohol construct and both of the latent variables. The fit indices result (Table 22) demonstrated an improvement in all fit indices when the cigarette frequency variable was used in alcohol and drug construct. Based on this result, the main model presented in the next section holds cigarette frequency for both of the latent variables.

TABLE 22 – Fit indices for confirmatory factor analysis

	Base model	Cigar with alc	Cigar in both
Robust CFI	0.924	0.914	0.931
Robust TLI	0.906	0.893	0.913
GFI	0.958	0.951	0.961
RMSR	0.058	0.072	0.045
RMSEA	0.110	0.118	0.106
Chi-square	806,791.72	919,953.36	739,288.49
BIC	22,496,266.82	22,609,428.46	22,428,876.12
AIC	22,491,640.35	22,604,801.99	22,424,157.11

Source: Elaborated by the author using PeNSE data.

At last, Table 23 presents the loadings for the SEM analysis. Each factor has the first estimate set to one, so every latent variable is accurately identified. The column Std.all returns the solution standardized and demonstrates that, with exception of the cigar variable, all loadings are bigger than 0.70; which is a good representation for each item. Additionally, the g factor represents the hierarchical latent variable that encounters the other constructs in one single broad risk factor. As one risk-taking is related to a propensity for other risky behaviors, this factor highly matters to our analysis. The next section presents the results from the main model, group models, and mediation analysis.

2.4 RESULTS

The main SEM model is presented in Table 24. Models 1-4 indicate the individual associations of the latent variables to schooling expectation. For all these cases the result was significant and negative, with higher magnitudes found in the g and sex-related factor. As the g construct is defined by all other constructs jointly, it is not possible to add all latent variables in the same model, once it would result in a linear combination.

[8] Root mean square of the residual. Good <0.06, acceptable 0.06-0.08, poor >0.01. See more (BUCHANAN, 2020)

TABLE 23 – Estimates for Latent Variables - Main model

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
rsex =						
agesex	1				1.169	0.87
preserv	0.539	0.005	118.793	0.0000	0.63	0.947
preserv1	0.549	0.005	116.856	0.0000	0.642	0.964
sexpartn	1.042	0.012	87.065	0.0000	1.218	0.774
drog =						
agedrug	1				0.976	0.85
mariju30	0.583	0.013	46.044	0.0000	0.569	0.98
crack30	0.299	0.005	54.636	0.0000	0.292	0.866
drug30	0.649	0.015	43.125	0.0000	0.633	0.956
cigar30	0.404	0.016	25.25	0.0000	0.394	0.461
alc =						
agealc	1				1.409	0.734
alc_freq30	0.806	0.0110	74.736	0.0000	1.136	0.892
alc_int30	1.162	0.016	72.323	0.0000	1.638	0.866
drunk	0.662	0.01	69.301	0.0000	0.934	0.839
probeb	0.419	0.006	74.376	0.0000	0.591	0.743
cigar30	0.169	0.006	27.174	0.0000	0.238	0.278
schooexp =						
maxschool	1				1.401	1
g =						
alc	1.097	0.016	69.477	0.0000	0.778	0.778
sexr	0.66	0.013	50.462	0.0000	0.565	0.565
drug30	0.564	0.014	39.655	0.0000	0.578	0.578

Source: Elaborated by the author using PeNSE data.

Consequently, Models 5-14 present the results for all possible compositions of 2 and 3 constructs.

The g factor remained negative and significant for all models except for the sex-related construct. Regarding alcohol, it was verified significance and positivity when regressed with g factor but was negative when regressed solely with drug construct. The sex-related construct remained negative, significantly, and with higher magnitudes for all models in which was included. Lastly, the drug latent variable repeated the same pattern observed in the alcohol latent, but with lower significance levels.

A common feature in the models presented in Table 24 is that when regressed with sex latent, all other variables lose significance. An interpretation may be that, as the association of sex-related construct is much stronger than the other latent variables when regressed with sex, the other associations became insignificant. Table 6 presents the correlation factor matrix[9] . For those related to the schooling expectation, sex-related latent variable correlated three times higher than alcohol and drugs, this characteristic shed light on the results found in models from Table 24.

To the multi-group models' application, some restrictions must be underlined. The implementation precludes that there is no high significant variance between groups in the loadings, intercepts, and residuals. This is the invariance assumption of the metric,

[9] The correlation matrix for the manifest variables is presented in Table 71, from Appendix

TABLE 24 – SEM Models

	g	Alcohol	Sex	Drugs
Model 1	-0.096***			
	(0.014)			
Model 2		-0.034***		
		(0.008)		
Model 3			-0.113***	
			(0.009)	
Model 4				-0.043***
				(0.014)
Model 5	-0.299***	0.135***		
	(0.037)	(0.021)		
Model 6	0.023		-0.125***	
	(0.019)		(0.012)	
Model 7	-0.122***			0.03*
	(0.018)			(0.018)
Model 8		0.014	-0.121***	
		(0.009)	(0.01)	
Model 9		-0.025***		-0.026*
		(0.009)		(0.015)
Model 10			-0.114***	0.003
			(0.01)	(0.015)
Model 11	-0.019	0.023	-0.116***	
	(0.061)	(0.029)	(0.017)	
Model 12	-0.595***	0.253***		0.146***
	(0.055)	(0.029)		(0.023)
Model 13	0.038		-0.128***	-0.015
	(0.038)		(0.013)	(0.018)
Model 14		0.015	-0.12***	-0.005
		(0.009)	(0.01)	(0.015)

Source: Elaborated by the author using PeNSE data.
Obs.: Cigar in alcohol and drug contructs.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 25 – Correlation factor matrix

	Sexual	Drugs	Alcohol	Expectation
Sexual	1			
Drugs	0.326	1		
Alcohol	0.439	0.45	1	
Expectation	-0.095	-0.029	-0.03	1

Source: Elaborated by the author using PeNSE data.

scalar, and strict steps, respectively. To a significant variance, it was considered a CFI[10] difference bigger than 0.01 (BUCHANAN, 2020). Therefore, to the multi-group model that presented significant variance, it was applied partial invariance, in which some manifest variables feature (loadings, intercepts, residuals) are released to 'gain' CFI levels and turn the difference bellow 0.01.

Three limitations must be addressed for these partial invariance models. First, it was not possible to apply sample weights[11]. Second, it would not allow the g factor. Three, cigar variable in both alcohol and drug constructs could not be implemented. The two last limitations are related to the fact that our model is non-congeneric, which means that a manifest variable is presented in more than one construct. Therefore, to the implementation of multi-group models, we followed the base model presented in the identification strategy subsection, which is a congeneric model.

Table 26 has the multi-groups results by gender. Boys generally had more significant coefficients and higher magnitudes compared to girls. A divergence regarding the main model is that for men, alcohol was also significant and negative when regressed with sex-related factor. For girls, the drug construct was significant and positive in model 7. Similar to the main model, sex-related construct remained the most reliable, but alcohol use for boys has a stronger association with schooling expectation variable when compared to girls.

TABLE 26 – SEM multi-group model by gender

	Men			Women		
	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs
Model 1	-0.050***			-0.020***		
	(0.005)			(0.005)		
Model 2		-0.088***			-0.077***	
		(0.005)			(0.006)	
Model 3			-0.042***			-0.014*
			(0.007)			(0.008)
Model 4	-0.011*	-0.083***		0.009	-0.082***	
	(0.006)	(0.006)		(0.005)	(0.007)	
Model 5	-0.046***		-0.011	-0.020***		0.000
	(0.06)		(0.008)	(0.005)		(0.009)
Model 6		-0.087***	-0.007		-0.083***	0.021
		(0.006)	(0.007)		(0.007)	(0.008)
Model 7	-0.011*	-0.083***	-0.002	0.004	-0.085***	0.018**
	(0.006)	(0.006)	(0.008)	(0.006)	(0.007)	(0.009)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

Table 27 has the results of the multi-groups by city. Alcohol had a stronger magnitude when regressed alone for capital cities, but more significant results were

[10] Comparative fit index reliability criteria that compares the fit of a target model to the fit of a null, model and it is not much sensitive to sample size. >0.95 good, >0.90 acceptable, <0.90 poor (HOOPER; COUGHLAN; MULLEN, 2008).

[11] To the best of our knowledge, unfortunately, 'lavaan.survey' package has not the possibility to apply weights in multi-groups with partial invariance.

found in the interior group, with positive results when regressed with sex-related factor. Further, for all significant results of the drug latent variable, the association was negative. While for interior cities the associations were weaker and with different signs. Maybe this result is due to the higher availability of the drug market in capital cities.

TABLE 27 – SEM multi-group model by area

	Capital			Interior		
	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs
Model 1	-0.042***			-0.027***		
	(0.005)			(0.005)		
Model 2		-0.109***			-0.101***	
		(0.006)			(0.006)	
Model 3			-0.052***			-0.026***
			(0.006)			(0.008)
Model 4	0.001	-0.110***		0.014**	-0.109***	
	(0.005)	(0.006)		(0.006)	(0.007)	
Model 5	-0.030***		-0.032***	-0.025***		-0.009
	(0.05)		(0.007)	(0.006)		(0.009)
Model 6		-0.106***	-0.010		-0.104***	0.014*
		(0.006)	(0.007)		(0.006)	(0.009)
Model 7	-0.005	-0.108***	-0.013*	0.012**	-0.109***	0.008
	(0.006)	(0.007)	(0.008)	(0.006)	(0.007)	(0.009)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

Multi-group application by type of school is presented in Table 28. The sex-related construct magnitudes are alike for both groups, but drug and alcohol magnitudes and significance were prevalently stronger for the private school group. All signs for the private schools' group were negative, while those significant in the public-school group were positive when regressed with any other latent variable. This could be related to higher absorption of risk in the schooling expectation of private school group, once their investment in education is higher compared to the public-school group.

Analyzing the multi-group model by region in Table 29. It is possible to see that alcohol associations were stronger in magnitude for the south region. The sex-related factor presented to be lower in magnitude in the north and northeast region compared to the mid-west, south, and southeast. Also, the only case in which alcohol and drug variables had significance when regressed with sex-related construct was for the south region, though not a strong relation all signs were negative. These results seem to demonstrate the existence of some regional differences among students. With an emphasis on the sex construct the middle-south part of Brazil might integrate risk behavior to the expectations at higher levels compared to north and northeast, inequality between regions is likely to be a source of this situation.

TABLE 28 – SEM multi-group model by school

	Public			Private		
	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs
Model 1	-0.023***			-0.038***		
	(0.004)			(0.006)		
Model 2		-0.076***			-0.088***	
		(0.005)			(0.009)	
Model 3			-0.019***			-0.057***
			(0.006)			(0.010)
Model 4	0.008*	-0.081***		-0.012*	-0.080***	
	(0.005)	(0.005)		(0.007)	(0.010)	
Model 5	-0.022***		-0.004	-0.027***		-0.037***
	(0.05)		(0.007)	(0.007)		(0.011)
Model 6		-0.079***	0.011*		-0.080***	-0.024**
		(0.005)	(0.006)		(0.009)	(0.010)
Model 7	0.006	-0.081***	0.008*	-0.007	-0.076***	-0.020*
	(0.005)	(0.005)	(0.007)	(0.007)	(0.010)	(0.011)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 29 – SEM multi-group model by region

	Midwest			South			Southeast			Northeast			North		
	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs
Model 1	-0.044***			-0.051***			-0.047***			-0.037***			-0.035***		
	(0.009)			(0.010)			(0.008)			(0.006)			(0.008)		
Model 2		-0.123***			-0.117***			-0.130***			-0.102***			-0.088***	
		(0.011)			(0.014)			(0.010)			(0.008)			(0.008)	
Model 3			-0.054***			-0.026**			-0.052***			-0.031***			-0.044***
			(0.011)			(0.012)			(0.011)			(0.011)			(0.011)
Model 4	0.001	-0.124***		-0.012	-0.110***		0.001	-0.130***		0.004	-0.104***		0.006	-0.091***	
	(0.010)	(0.012)		(0.012)	(0.016)		(0.009)	(0.012)		(0.007)	(0.008)		(0.009)	(0.009)	
Model 5	-0.032***		-0.033**	-0.054***		0.006	-0.037***		-0.026**	-0.035***		-0.005	-0.026***		-0.027**
	(0.010)		(0.013)	(0.012)		(0.014)	(0.009)		(0.013)	(0.007)		(0.012)	(0.009)		(0.013)
Model 6		-0.120***	-0.011		-0.125***	0.017		-0.129***	-0.001		-0.105***	0.017		-0.086***	-0.010
		(0.012)	(0.012)		(0.015)	(0.013)		(0.011)	(0.012)		(0.008)	(0.012)		(0.009)	(0.012)
Model 7	0.005	-0.122***	-0.014	-0.021*	-0.116***	-0.027*	0.002	-0.130***	-0.002	0.001	-0.106***	0.016	0.011	-0.090***	-0.016
	(0.011)	(0.013)	(0.013)	(0.013)	(0.016)	(0.014)	(0.010)	(0.012)	(0.013)	(0.008)	(0.009)	(0.012)	(0.009)	(0.009)	(0.013)

Source: Elaborated by the author using PeNSE data.

Note.: *** p<0.01. ** p<0.05. * p<0.1.

Multi-group models by color/race are presented in Table 30. The black and brown/mixed skin group follow the same pattern of significance and signs from the main model. As black and brown/mixed skin represent a major part of the population, it is reasonable that the results from these group models assimilate to the main model. A higher magnitude to sex-related factor was found for the white group and positive associations were verified for drug construct related to the yellow group. Also, the indigenous group had a significant and positive association for alcohol variables when regressed with sex-related factor, which is a divergent result compared to the main model.

TABLE 30 – SEM multi-group model by color/race

	Indigenous			Yellow			Black			Brown/Mixed			White		
	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs	Alcohol	Sexual	Drugs
Model 1	0.011			-0.031**			-0.041***			-0.035***			-0.030***		
	(0.020)			(0.017)			(0.011)			(0.005)			(0.006)		
Model 2		-0.086***			-0.086***			-0.099***			-0.096***			-0.105***	
		(0.022)			(0.020)			(0.012)			(0.006)			(0.007)	
Model 3			-0.037			0.025			-0.035**			-0.039***			-0.030***
			(0.028)			(0.025)			(0.015)			(0.008)			(0.009)
Model 4	0.057***	-0.115***		0.002	-0.087***		-0.002	-0.098***		0.002	-0.097***		0.012*	-0.112***	
	(0.022)	(0.025)		(0.019)	(0.022)		(0.012)	(0.014)		(0.006)	(0.007)		(0.007)	(0.008)	
Model 5	0.032		-0.060*	-0.049**		0.058**	-0.038***		-0.010**	-0.029***		-0.019**	-0.027***		-0.011
	(0.023)		(0.032)	(0.019)		(0.028)	(0.012)		(0.017)	(0.006)		(0.009)	(0.007)		(0.010)
Model 6		-0.085***	-0.004		-0.106***	0.021***		-0.100***	0.002		-0.096***	-0.002		-0.109***	0.014
		(0.024)	(0.030)		(0.021)	(0.027)		(0.013)	(0.016)		(0.007)	(0.008)		(0.008)	(0.009)
Model 7	0.070***	-0.111***	-0.043	-0.018	-0.099***	0.082***	-0.004	-0.098***	0.004	0.003	-0.097***	-0.004	0.009	-0.113***	0.009
	(0.024)	(0.025)	(0.033)	(0.020)	(0.023)	(0.029)	(0.013)	(0.014)	(0.017)	(0.006)	(0.007)	(0.009)	(0.007)	(0.008)	(0.010)

Source: Elaborated by the author using PeNSE data.

Note.: *** p<0.01. ** p<0.05. * p<0.1.

The loadings for each multi-group model are presented in Tables 72 - 79. The multi-groups by city and color/race were set to the configural estimation[12] (Table 74), which means that no changes in invariance were necessary. The multi-group by gender had four variables released with variance released, they were the number of partners, age of sex onset, crack frequency, and preservative use. The cause for differences is probably in the overreport of sex relation by boys compared to girls, chapter 1 raised this discussion and pointed out that this feature is commonly found in the literature (WELLINGS et al., 2006). Likewise, the crack frequency had to be released in the multi-group by gender because male students had a bigger variance.

For region multi-group crack, cigarette frequency, and age of drug initiation had to be released. It occurs because higher variance was verified for the south and southeast groups compared to other regions. At last, multi-groups by school type had preservative and number of partners, manifest variables released, once public schools have significantly higher variance in these variables compared to the private school group.

The mediation effects are an important analysis as they describe how other variables not stated in the main measurement model can indirectly influence the schooling expectation outcome. Figure 2 contextualizes the mechanism of mediation. First, we make a regression of the risk related to the mediator (a), after the mediator is regressed against the response variable of schooling expectation (b). Multiplying (a) to (b) we have the mediation (indirect) effect, (c) represents the direct effect that, in sum with (a) times (b) returns the total effect. The mediations were tested considering each construct alone to avoid the high influence that sex-related factor has on the other variables.

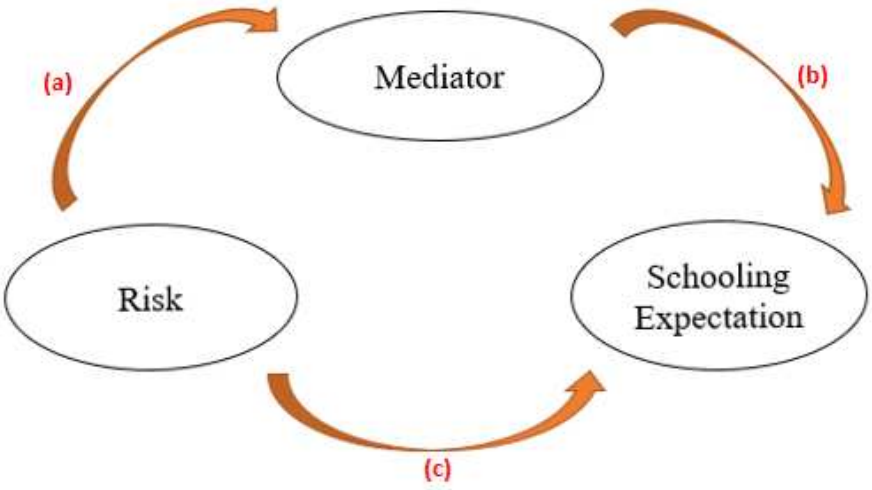

FIGURE 9 – Mediation Effect

Twelve mediation effects were tested, of those two represent constructs (bullying, peer effects) and ten observed variables. For the bullying construct, the following

[12] Configural estimation is the case where no partial invariance was required.

variables were considered: they are felt alone in the recent 12 months, colleagues treat well, being victimized in the recent 30 days, suffered bullying at least once in life. The peer effects construct considers if the student has friends that drink alcohol and use drugs. The loadings for these estimations are stated in Table 80 and 81, from Appendix. For the other cases, the loadings were not inserted, once the observed variables did not highly change the composition of the constructs.

Bullying mediation effect is presented in Table 31; for all risk variables bullying had a positive and significant effect. When bullying mediation was regressed, related to schooling expectation, there was a reversal of signs, being negative and significant. The indirect effect of bullying on risky behaviors ends up being negative, as risk is positively related to bullying construct and the last one is negatively related to schooling expectation. The total effect is increased by the mediation effect, that negatively associates with the educational prospect of the students.

TABLE 31 – Bullying mediation on risk behavior

	Bullying	Expectation		Effect	
	Risk	Bullying	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.021***	-0.616***	-0.067***	-0.013***	-0.080***
	(0.001)	(0.062)	(0.006)	(0.001)	(0.006)
Alcohol	0.012***	-0.684***	-0.027***	-0.008***	-0.035***
	(0.001)	(0.063)	(0.004)	(0.001)	(0.004)
Sexual	0.006***	-0.801***	-0.102***	-0.005***	-0.107***
	(0.000)	(0.067)	(0.004)	(0.001)	(0.004)
Drugs	0.006***	-0.834***	-0.029***	-0.005***	-0.034***
	(0.001)	(0.067)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

Regarding peer effect mediation, Table 32 indicates a positive association with peer effects for all risks analyzed. But the association of peers to schooling expectation has some specificities. For the g factor it was positive, while for alcohol and drug factors negative, with no association found for sex-related latent variable. In such context, the mediation was positive considering overall risk (g), and negative when alcohol and drug construct were analyzed.

Age mediation on schooling expectations (Table 33) indicated a positive association to the risk and negative to schooling expectation. PeNSE data is based on students from ninth grade, hence being older is likely to mean a school delay. Also, as they are older, they are more likely to have experienced risky behaviors, as showed in Table 33. Thus, the indirect effect is negative for all risk behaviors, and the total effect remains solely to alcohol and sex-related construct, once the direct effect of g and drug were not significant.

The number of people at home (Table 34) demonstrated to be negatively correlated with risky behaviors, except sexual. Also, this mediator is negatively correlated with

TABLE 32 – Peer effect mediation on risk behavior

	Peer effect	Expectation		Effect	
	Risk	Peer effect	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.549***	0.129***	-0.160***	0.071***	-0.089***
	(0.004)	(0.030)	(0.019)	(0.017)	(0.006)
Alcohol	0.291***	-0.035**	-0.026***	-0.010**	-0.036***
	(0.003)	(0.017)	(0.006)	(0.005)	(0.004)
Sexual	0.236***	-0.000	-0.107***	-0.000	-0.107***
	(0.003)	(0.011)	(0.005)	(0.003)	(0.004)
Drugs	0.422***	-0.064***	-0.007	-0.027***	-0.034***
	(0.004)	(0.008)	(0.006)	(0.003)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 33 – Age mediation on risk behavior

	Age	Expectation		Effect	
	Risk	Age	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.324***	-0.236***	-0.006	-0.076***	-0.082***
	(0.005)	(0.005)	(0.006)	(0.002)	(0.006)
Alcohol	0.174***	-0.240***	0.006*	-0.042***	-0.035***
	(0.003)	(0.005)	(0.004)	(0.001)	(0.004)
Sexual	0.296***	-0.222***	-0.041***	-0.066***	-0.107***
	(0.003)	(0.005)	(0.005)	(0.002)	(0.004)
Drugs	0.157***	-0.238***	0.003	-0.037***	-0.034***
	(0.004)	(0.005)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

expectations. To the extent that two negative correlations result in a positive indirect effect due to the multiplication, we need to change our understanding regarding this case and assume these effects as negative. As a result, we have a negative indirect effect on schooling expectation for all risky behaviors mediated by the number of people at home.

Mediation effect for internet is presented in Table 35. There is a positive association of risky behaviors, except for sex-related factor. From now on, considering the regression of the mediator on schooling expectation, once again, the internet is associated with higher expectations. Therefore, the indirect effects of risky behaviors are mostly negative. One possible interpretation for this result is that students with internet access are more prone to keep contact and establish relations via the internet, being integrated into groups that allow some risky behaviors experimentation.

The same pattern followed the mediation effects of someone in the family having a car and having a housekeeper (Table 36 - 37). Although with different magnitudes

TABLE 34 – N household mediation on risk behavior

	N household	Expectation		Effect	
	Risk	N household	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	-0.021***	-0.097***	-0.083***	0.002***	-0.080***
	(0.006)	(0.003)	(0.006)	(0.001)	(0.006)
Alcohol	-0.028***	-0.097***	-0.038***	0.003***	-0.035***
	(0.004)	(0.003)	(0.004)	(0.000)	(0.004)
Sexual	0.036***	-0.094***	-0.104***	-0.003***	-0.107***
	(0.005)	(0.003)	(0.004)	(0.000)	(0.004)
Drugs	-0.026***	-0.097***	-0.037***	0.003***	-0.034***
	(0.006)	(0.003)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 35 – Internet mediation on risk behavior

	Internet	Expectation		Effect	
	Risk	Internet	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.017***	0.592***	-0.090***	0.010***	-0.080***
	(0.002)	(0.013)	(0.006)	(0.001)	(0.006)
Alcohol	0.016***	0.591***	-0.045***	0.009***	-0.035***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Sexual	-0.017***	0.569***	-0.097***	-0.010***	-0.107***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Drugs	0.012***	0.585***	-0.041***	0.007***	-0.034***
	(0.001)	(0.013)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

significance and signs were the same for the three tables[13] . As these variables related to the students' economic aspects, it would have been better to implement a formative construct within the three. Unfortunately, the package software used to develop the model only accepts reflexive constructs. Despite that, the same nature of results being found for the three mediation estimations reinforces that economic aspects can be associated with positivity in alcohol and drug risky behaviors, but with negativity association with sex-related latent variable.

Smoking parents mediation effects are presented in Table 38. Risky behaviors correlated positively with this mediator. At the same time, when regressed with schooling expectation, smoking parents variable association was negative. This brings an indirect effect for all risky behavior constructs when parents' addiction to cigarettes is considered. Such an element is important, once indicates that parents might have an influence on the risk behavior developed by their children.

Missed classes without parents' permission had a negative mediation effect for

[13] Except for g factor being insignificantly for car mediation.

TABLE 36 – Car mediation on risk behavior

	Car	Expectation		Effect	
	Risk	Car	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.001	0.403***	-0.081***	0.000	-0.080***
	(0.002)	(0.010)	(0.006)	(0.001)	(0.006)
Alcohol	0.009***	0.406***	-0.039***	0.004***	-0.035***
	(0.001)	(0.010)	(0.004)	(0.001)	(0.004)
Sexual	-0.030***	0.387***	-0.096***	-0.012***	-0.107***
	(0.002)	(0.010)	(0.004)	(0.001)	(0.004)
Drugs	0.008***	0.404***	-0.037***	0.003***	-0.034***
	(0.002)	(0.010)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 37 – Housekeeper mediation on risk behavior

	Housekeeper	Expectation		Effect	
	Risk	Housekeeper	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.011***	0.339***	-0.084***	0.004***	-0.080***
	(0.001)	(0.014)	(0.006)	(0.001)	(0.006)
Alcohol	0.009***	0.337***	-0.039***	0.003***	-0.035***
	(0.001)	(0.014)	(0.004)	(0.000)	(0.004)
Sexual	-0.004***	0.325***	-0.106***	-0.001***	-0.107***
	(0.001)	(0.014)	(0.004)	(0.000)	(0.004)
Drugs	0.007***	0.333***	-0.036***	0.002***	-0.034***
	(0.001)	(0.014)	(0.005)	(0.000)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 38 – Parents smoke mediation on risk behavior

	Parent smoke	Expectation		Effect	
	Risk	Parent smoke	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.072***	-0.133***	-0.071***	-0.010***	-0.081***
	(0.002)	(0.012)	(0.006)	(0.001)	(0.006)
Alcohol	0.042***	-0.148***	-0.029***	-0.006***	-0.035***
	(0.001)	(0.012)	(0.004)	(0.001)	(0.004)
Sexual	0.035***	-0.134***	-0.102***	-0.005***	-0.107***
	(0.001)	(0.012)	(0.004)	(0.000)	(0.004)
Drugs	0.036***	-0.157***	-0.028***	-0.006***	-0.034***
	(0.002)	(0.012)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

all risky behaviors (Table 39). The risk constructs associated positively with missing classes, while missing classes associated negatively with schooling expectation. The same pattern followed the bully victimizer mediation effect (Table 40), with the difference that for g construct, the regression of bully victimizer on schooling expectations was not significant. As missing classes represents an apathy to educational achievements, it seems reasonable to carry such results. Also, bully victimizers usually present a more risk-taking and aggressive behavior, which could affect a propensity to risky behaviors and unconcern with schooling prospect.

TABLE 39 – Missing classes mediation on risk behavior

	Miss classes	Expectation		Effect	
	Risk	Miss classes	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.105***	-0.262***	-0.054***	-0.027***	-0.081***
	(0.002)	(0.013)	(0.006)	(0.001)	(0.006)
Alcohol	0.057***	-0.282***	-0.019***	-0.016***	-0.035***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Sexual	0.061***	-0.248***	-0.092***	-0.015***	-0.107***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Drugs	0.062***	-0.291***	-0.016***	-0.018***	-0.034***
	(0.002)	(0.013)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

TABLE 40 – Bully victimizer mediation on risk behavior

	Bully	Expectation		Effect	
	Risk	Bully	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	0.115***	-0.022	-0.079***	-0.003	-0.081***
	(0.002)	(0.014)	(0.006)	(0.002)	(0.006)
Alcohol	0.066***	-0.053***	-0.032***	-0.003***	-0.035***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Sexual	0.054***	-0.029**	-0.106***	-0.002**	-0.107***
	(0.001)	(0.013)	(0.004)	(0.001)	(0.004)
Drugs	0.063***	-0.071***	-0.030***	-0.004***	-0.034***
	(0.002)	(0.013)	(0.006)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

We also tested the mediation effect of comprehensive parents on schooling expectations (Table 41). In this case, risky behaviors were negatively associated with comprehensive parents. At the same time, parents' comprehension was positively associated with schooling expectations. To this mediation, the understanding of the indirect effect must be changed and considered as positive, although it shows as negative. This mediation result might reflect that environments, where students feel more accepted, might help to avoid the practice of risky behaviors.

TABLE 41 – Comprehensive parents mediation on risk behavior

	CompParent	Expectation		Effect	
	Risk	CompParent	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	-0.093***	0.018*	-0.079***	-0.002*	-0.081***
	(0.002)	(0.010)	(0.006)	(0.001)	(0.006)
Alcohol	-0.055***	0.032***	-0.034***	-0.002***	-0.035***
	(0.001)	(0.010)	(0.004)	(0.001)	(0.004)
Sexual	-0.031***	0.030***	-0.106***	-0.001***	-0.107***
	(0.002)	(0.010)	(0.004)	(0.000)	(0.004)
Drugs	-0.042***	0.043***	-0.032***	-0.002***	-0.034***
	(0.002)	(0.010)	(0.005)	(0.000)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

Lastly, HPV vaccine mediation is analyzed (Table 42). The risk constructs were associated negatively with the HPV vaccine, while the regression on schooling expectation by the mediator was positive and significant for all latent variables. A common sense among the population is that the HPV vaccine anticipates the sexual onset (DO; WONG, 2012). Although we cannot affirm that these results are not true, we can say that it is not positively associated with sexual risk behavior. More, we can also say that is negatively associated with other risky behaviors.

TABLE 42 – HPV vaccine mediation on risk behavior

	HPV	Expectation		Effect	
	Risk	HPV	Risk	Indirect	Total
	(a)	(b)	(c)	(a)*(b)	(a)*(b)+(c)
g	-0.047***	0.280***	-0.068***	-0.013***	-0.081***
	(0.002)	(0.010)	(0.006)	(0.001)	(0.006)
Alcohol	-0.019***	0.289***	-0.030***	-0.005***	-0.035***
	(0.001)	(0.010)	(0.004)	(0.000)	(0.004)
Sexual	-0.095***	0.247***	-0.084***	-0.023***	-0.107***
	(0.001)	(0.010)	(0.005)	(0.001)	(0.004)
Drugs	-0.031***	0.291***	-0.025***	-0.009***	-0.034***
	(0.002)	(0.010)	(0.005)	(0.001)	(0.005)

Source: Elaborated by the author using PeNSE data.
Note.: *** p<0.01. ** p<0.05. * p<0.1.

To robustness check, fit indices were verified. They are CFI, TLI[14] , GFI[15] , RMSR, RMSEA[16] , Chi-square BIC, and AIC. To the cases that were applied sample weights, the fit indices will return robust versions for CFI and TLI, once for those de ML estimator had robust standard errors application. For the multi-group, the indices of CFI and TLI

[14] Tucker Lewis Index. >0.95 good, >0.90 acceptable, <0.90 poor.

[15] Proportion of variance accounted for by the estimated population covariance (Goodness of fit) >0.95 good, >0.90 acceptable, <0.90 poor. See more Hooper, Coughlan e Mullen (2008)

[16] Root Mean Square Error of Approximation. <0.05 good, <0.01 acceptable, >0.01 poor. See more (BUCHANAN, 2020)

are the regular ones. Table 82 from the appendix has the result of robustness check for the main model. CFI index is above 0.93 for all models, TLI had a minimum of 0.913, and GFI of 0.960. For the residual indices, RMSR is below 0.047 for all models, but RMSEA is around 0.105-0.106.

The multi-group models had a decline in the fit qualities but remained acceptable for most of them (Tables 83 - 87). In the multi-group by gender (Table 33) TLI was below 0.9, which indicates a poor quality for this index. Also, for all multigroup models RMSEA was above 0.01. The other fit indices were at acceptable levels. The mediation models had better indices (Tables 88 - 89). Although in some cases GFI was below 0.90 (peer effects and age), all CFIs and TLIs were at acceptable levels. Moreover, for this analysis, all RMSR and RMSEA were below 0.01, which indicates acceptable levels for all models applied.

Still, our fit indices are not ideal in terms of goodness of fit, usually, notorious levels for reliability criteria are above 0.95 (HAIR et al., 2009). Also, our RMSEA results for the main and multi-group models were poor. Accordingly, as these results require precaution, our analysis concerning them has been implemented focusing on the associations and correlations, with no causal inference being concluded. The next section will discuss the results and analyze interpretations based on the theory underlined in the precedent sections.

2.5 DISCUSSION

The result section presented the models of risk behavior related to schooling expectation applying structural equation models to a relevant set of Brazilian students. Considering the behavioral economic approach implemented in the literature review some discussion must be addressed regarding prior results. A negative association was usually identified between some risky behaviors and schooling expectations, especially for risky sexual behavior. On the other hand, there were positive associations and variance in the results for alcohol and drug constructs.

Our results demonstrated that the sex construct was negatively and significantly related to the schooling expectation in almost all estimates. A sophistication level can be interpreted from this result, implying that students' perception of this behavior represents a factor that might reduce expectations. Although the data do not allow direct observation of this relationship, our analysis relies on the dramatic effects that an unwanted pregnancy would have on the lives of youths. That is, unlike other constructs, the sex construct has implications and costs that are more clearly perceived by these students and may be better incorporated into expectations, while for other activities, long-term costs are not perceived as relevant.

Moreover, the g factor presented some variance in the results, being positive or insignificant when regressed with a sex-related factor or negative when analyzed with other constructs. To this extent, it is important to recall that the practice of many risk

behaviors jointly might perceive worse consequences of bad behavior. For instance, Do et al. (2020) verified that individuals drinking alcohol and using other stimulants before sex had a higher likelihood of unintended pregnancy.

For the alcohol construct, the results were positive mostly when regression considered the general and sexual construct. The reversal of signs can be related to a weak association between schooling expectation and risk. Also, it can be interpreted that the association between the sexual construct and the schooling expectation is so strong that the other relations become meaningless. Recovering the theoretical literature, this might also imply a degree of naivety in students' perceptions of risky, once they might be seen as negative in general, but not as strong as other risks.

Even though students are not allowed legally to make use of alcoholic drinks or any other substance, it is common for adolescents to have such experiences before maturity. Such aspects may be related to alcohol consumption be more socially acceptable. As stated by Becker e Murphy (1988), goods that are highly addictive to most people tend to have a bimodal distribution of consumption, with one mode located near abstention.

Still, strong drugs could have such distribution, but alcohol consumption is likely to be more continuous, once alcoholic beverages are not perceived as addictive for many people. Besides, early initiation of drinking warrants attention to a larger externalizing behavior pattern, with poor self-control, disinhibition, and reward-seeking behavior problems. Those situations might be associated with greater alcohol consumption, as well as an increased risk of continued drinking in early adolescence (ISAKSSON et al., 2020).

The same pattern of results followed the substance use construct, but this one had more significance and sign variability. This may be seen as a counterintuitive result. However, it is worth emphasizing that the manifest variables used for the drug construct relate to the use of nicotine, marijuana, crack and the age he got the first illicit substance, which can be of any type. Except for crack, the other substances that are captured by the construct may not appear to the students costly enough for a high perception of risk and a negative relation with the schooling expectation. Thus, a large part of the manifest variables may be understood by the students as not offering potential risk.

The multi-group estimations followed the same pattern of the main model with some differences being noted between groups. First, the alcohol construct was stronger in magnitude for males, as well as for interior cities. Private schools registered more significance and higher magnitudes to risky behavior compared to public schools. South region had stronger magnitudes for alcohol and the North had a lower coefficient to sex latent variable. Lastly, multi-group by color demonstrated that the results for the main model assimilates to the ones observed in the black and brown/mixed skin groups.

Important aspects are revealed when the mediator variables are analyzed. The first case was bullying, which appeared to be a promotive factor in reducing expectations.

Also, for students that practice bullying the results were much the same, but with stronger magnitudes. The victim and victimizer on bullying dynamic represent two opposites, one has a propensity to internalizing behaviors, while the other has to externalize behaviors. Internalizing behaviors are those associated with being withdrawn or struggling to interact and participate in normative ways. In contrast, externalizing behaviors are those outward-directed, aggressive, delinquent, or disruptive (MIHALEC-ADKINS; COOLEY, 2020). Still, both can affect risk behavior once they might work as an escape attribute, and be negatively correlated with schooling expectation as the school becomes an unpleasant environment.

Peer effect mediation had a dubious result, being associate with higher risk behavior, but also with higher expectations. This result is related to the fact that although risk peers influence risk, they also stimulate the student to feel belonged and accepted in the environment. This may have o promotive impact on raising expectations. Also, as observed in other studies, the presence of peers might increase the salience of the immediate rewards of a risky choice, being possible to transform choices of late adolescents into ones that are similar to less mature teenagers (O'BRIEN et al., 2011). As our dataset has students of mean age between 14 and 15 such discounting can have deepened impacts.

Students' age is a relevant variable once older students probably have an educational delay and might be prone to have a lower expectation. Related to the number of people in the household, an interpretation is that this variable captures a group of vulnerable students that probably had to enter the labor market and is likely to already have an "adult life". Although very impressionistic, this could probably explain why the other risk variables were negatively correlated with the mediator while sex was positive.

Internet, car, and housekeeper mediators aimed to verify how economic aspects could be associated with risk and mediate schooling expectations. These variables had the same pattern and allowed the interpretation that students with more economic access are prone to have a negative association with sex-related factor, but correlates positively with other risky behaviors. It seems that this group acknowledge the negative consequences of sex-related factor, but are willing to engage on other risk behaviors. Young people that have well-structured families, with high financial and social capital might also have a higher prospect of human capital. So, they possibly address higher costs of succumbing to some harmful behaviors and might be more willing to accept the delayed benefits of higher education (COLEMAN, 1988; BECKER; MURPHY, 1988; LOEWENSTEIN; PRELEC, 1992).

The mediation regarding parents indicated that nicotine addiction was associated with a negative effect on the schooling expectation of the children. At the same time, having comprehensive parents had a positive association with schooling prospects and a negative with risky behaviors. Such results support the view that an environment of acceptance and without references to bad behavior can contribute to avoiding student engagement on risk helping to raise expectations.

As familiarity and involvement are components to the analysis of risk as controllable, the more a person engages in risk, the more he/she becomes willing to believe that the risk would not have much consequence to their lives (LANGER, 1975). The contact with people that make use of these goods might also create an illusion of control perspective. Such an aspect is important because people do not consider risk issues in isolation from other social issues or the views of their peers and family. People validate perceptions, share cultural values, and are also able to amplify or attenuate information (KASPERSON et al., 2000).

Moreover, missing classes had a negative mediation impact, which is something expected. As students that miss classes do not demonstrate much preoccupation with the content lost, they are likely to have lower expectations from school. Regarding HPV vaccine mediation, it had a negative association with all risk constructs and appeared positively associated with schooling expectation. As people that got the HPV vaccine demonstrate some level of prudence in this act, maybe they give more relevance to the future outcomes and expect returns from a higher schooling expectation.

Alternatively, as stated by Almeida e Araújo-Júnior (2016), the discount rate on the future value of health can also be held as undefined, once children do not have enough maturity and knowledge to formulate expectations related to intertemporal maximization of their health utility. Indeed, such analysis seems to have an echo for the alcohol and drug constructs, once their results underlined a high prevalence of naivete. On the other hand, for sex-related factor, for example, we have another understanding, because even for 14-15 years-old children such behaviors can be recognized as highly impacting their future.

Although the discount utility theory is used solely with interpretation objective, our results seem to supports the understanding that, at some level, risky behaviors can influence intertemporal analysis. In such context, sex-related construct, e.g., can highly correlate negatively with schooling prospect.

The relation between engagement with sexual risk had been highly sophisticated. Considering the analysis of preferences made in the literature review, it is interpreted that individuals are estimating $\hat{\beta} = \beta$[17] for such behavior. Although the literature points that there is a gap between intention and practice related to sexual risk (DO et al., 2020), the consequences of early pregnancy for the future schooling of young people can be classified as dramatic, especially in the case of women (FRANÇA; FRIO, 2018). In this sense, public policies aiming to raise awareness and activities that make young people experience a clearer analysis of the effects of early pregnancy and STD infection could be very important.

The behaviors of alcoholic risk and substances, presented in most of the estimates with high variation, suggest a naive interpretation of how such risks may be

[17] $\hat{\beta} = \beta$ means that the person is sophisticate and correctly predicts future behavior

associated with the schooling expectation. This suggests a $\hat{\beta} < 1$[18] for such risky behaviors. For that reason, public policies could help with awareness regarding such engagements, to make evident the costs of risky activities on long-term well-being, as well as to make the perception of these young people about such activities more connected with other aspects of their lives, such as education, employment, productivity, and health.

In general, these questions shed light on some aspects of young people's behavior, showing that, depending on the type of risk and the individual's characteristics, the answer about schooling expectations can change, and can be understood by perceptions of sophistication and naivete, which may suggest directions to public policies. As we assume that present-biased preferences have implications on welfare, the welfare loss analysis in terms of long-run utility is the appropriate criterion to analyze the welfare loss, as proposed by O'Donoghue e Rabin (1999).

Although it is not possible to directly observe the welfare loss, the study demonstrates a negative relation between risk behavior and a higher schooling prospect for the sex-related construct, which is aligned to the theoretical model's interpretation. As stated by O'Donoghue e Rabin (2001), a problem in the youth's decision is likely to recognize that current indulgence might lead to future indulgence. As a result, policymakers could create incentives to encourage youths to take into account the linkages between decisions.

To acknowledge practical implications incentives matter. Those incentives may be focused on the student's awareness or in the policies to avoid access to risky behaviors. As sophistication might help when one wants to quit an addiction, working on nudge policies to make people more sophisticated could prevent the development of addiction, as long as the target is not driven by the pessimism effect. On the other hand, a condescendence with naivete on youth could be harmful, as they may involve in some addiction and may repeatedly delay quitting believing that it is possible to make it later. As a result, a welfare loss is likely to happen.

If the student is likely to be sophisticated, the focus on commitment devices to help with self-control problems is the main concern. If people are naïve, policies aiming at educating people about the loss of control, or providing incentives to the use of commitment devices might have some impact, to make them more sophisticated (FREDERICK; LOEWENSTEIN; O'DONOGHUE, 2004). One aspect raised by O'Donoghue e Rabin (2001) is the appealing feature of these incentives related to the little harm that youths who are rationally engaging in risk behaviors suffer. Therefore, raising perceptions of the severity of bad outcomes might have some effect on student's risk-taking without affecting the rational addicted. For instance, the author cites examples as exaggerating the costs of being pregnant, acquiring AIDS, or being addicted, among others.

[18] $\hat{\beta} < 1$ means that the person is partially naive and incorrectly predicts future behavior.

Accordingly, one path in public policy aiming at self-control mechanisms is to make information more accessible and more available, with public policies devising to teach and disseminate effective self-control (HERRNSTEIN; PRELEC, 1992). The personal rule approach of (AINSLIE; HASLAM, 1992), is structured to make people more sophisticated. It is based on the understanding that personal rules implementation is a learnable skill, which is structured in a recursive (not tautological) manner. In general, social skills are learned or socialized abilities that allow individuals to engage in appropriate behaviors and interactions, refraining from behaviors that elicit negative reactions from others (MIHALEC-ADKINS; COOLEY, 2020). Thus, this mechanism might work to show people how their decisions affect future welfare. As adolescence is a phase that the absorption of information is expected, this target represents a suitable approach for such implementation.

In contrast, our results could also support Steinberg (2004) view for alcohol construct for example. His analysis considers risk-taking during adolescence as likely to be normative, biologically driven, and inevitable. Therefore, little can be done concerning intervention that might attenuate reward sensitivity or accelerate the maturation of self-regulatory competence. To the author, a more profitable strategy might focus on limiting opportunities for immature judgment to have harmful consequences. Additionally, it is worth mentioning how addicted people respond to variations in price. Becker e Murphy (1988) show that the long-run demand for addictive goods tends to be more elastic than the demand for nonaddictive goods. Dealing with young people, the consequences in the demand are heightened, once they usually have a low-income level or no income. In such context, strategies such as raising the price of cigarettes, more vigilantly enforcing laws governing the sale of alcohol, expanding access to mental health and contraceptive services would likely be more effective in limiting adolescent risky behaviors than strategies aimed at making adolescents wiser, less impulsive, or less short-sighted (STEINBERG, 2004).

In this paper, we believe that both aspects can be presented in the strategies implemented by policymakers. Interventions must address social competence, with decision-making and self-esteem features, as well as social influence to deal with peer pressure and high-risk situations (SINGH et al., 2020). Understanding what impedes individuals from taking the desired action helps inform a productive margin along which it is possible to apply a policy intervention (MADRIAN, 2014). Norms, sanctions, and effective policies to avoid access to these behaviors are critical. However, our results support the belief that students can be treated as decision-makers capable of making analyses and decide on complex choices. Helping them to recognize suboptimal choices and develop self-regulators could bring valuable skills to this population.

It is worth mentioning that the study presented has limitations. First, PeNSE data is self-reported, so that there may be several biases in the responses. Second, our analysis was not able to calculate discount rates, despite the theoretical discussion. Three, the multiple-selves model can be a source of critics regarding its metaphorical

context. Still, the paper made an extensive literature review and applied a structural equation model to a relevant set of the Brazilian student population. The missing aspects can be a source of future research.

2.6 CONCLUSION

This study sought to analyze risk behavior related to the schooling expectation based on behavioral economic models. With the application of structural equation models, a negative association was usually identified for sexual risk behavior related to schooling expectation. On the other hand, there were positive associations, in some cases, for alcohol and substance use. Multi-group models and mediation effects were tested to verify outside influence on risk behavior and schooling prospect. The main interpretation is that policies focusing on developing social skills and self-control are relevant departures for students' analysis of the decision process. Also, the application of policies aimed to impede youth access to risky behavior is likely to have a strong impact on their risk engagement, as well as on their schooling expectation.

3 AFFECTIVE RISK ANALYSIS AND THE IMPACT OF BULLYING ON THE RISK BEHAVIOR OF YOUNG PEOPLE

3.1 INTRODUCTION

Risk behavior can be defined as activities potentially harmful for physical or mental health such as tobacco, alcohol, illicit substances, and unprotected sex with possible implications for the youth's well-being (GRUBER, 2001; ALMEIDA; ARAÚJO-JÚNIOR, 2016). Adolescents may be vulnerable to risk engagement during adolescence, which makes the study of determinant aspects for greater involvement in risky activities relevant. Bullying is an aspect associated with a higher prevalence of risky behaviors, with students who practice or suffer from it (MALTA et al., 2019; LEE et al., 2020). Thus, bullying might have intrinsic characteristics that cause the decision-making process for risk engagement to more chronic among victims and victimizers, when compared to individuals who are not related to such practices.

Bullying, in its classic definition, is characterized by the practice of offense, verbal or physical, with the intention of cause harm to the victim in a process that exists an imbalance of power and occurs repeatedly over time (OLWEUS, 1993). Such an experience may have implications that reflect poor academic performance, reduction of classroom participation, and school attendance, as students become less engaged to these activities (CORNELL et al., 2013). Bullying can also have more serious consequences, as the development of severe psychological problems, such as self-mutilation (LI et al., 2020). It is worth mentioning that students who practice bullying are more likely to be involved in risky behaviors, feel alone, and have episodes of violence at home. This occurs because aggressors also suffer the consequences of the phenomenon and help to maintain it in the school environment (MELLO et al., 2017).

Behavioral economics studies the decision-making processes based on a positive perspective that aims to describe how individuals choose. Research that analyzes adolescents' risk behaviors in economics usually focuses on the role of cognition. In contrast, affective quality of imagery can play an important role in adolescent risk-taking stimulating emotional response to the implementation of the decision, turning into a strong predictor of health-threatening and health-enhancing behaviors (KASPERSON et al., 2000; LOEWENSTEIN, 2004). This effect is a demarcation of positive or negative quality of stimulus experienced as a feeling state in which a person can respond to a decision considering the emotional marker related to the risk (SLOVIC et al., 2007). Equally, such perspective underlies people do not consider risk in isolation from other social structures, such as their peers' views (KASPERSON et al., 2000).

Therefore, behavioral economics offers concepts that may help to describe young people's involvement in risk engagement related to the victimization phenomenon.

In this study, the affect heuristic[1] is considered to analyze student's decisions, in which judgments and affective states can alter the risk perception (SLOVIC et al., 2002). Moreover, as friends and colleagues can provide reference points for perception validation, we sought to verify the impact of bullying on risk behavior assessment considering the behavioral concepts to establish the decision-making process.

Based on PeNSE data, the method of Propensity Score Matching (PSM) with closest neighbor technique is carried out to create a counterfactual and provide comparability between treatment and control groups. The advantage of PSM is verified in the control of the observable characteristics, which in the fulfillment of its assumptions allows inference of causality (ROSENBAUM; RUBIN, 1983). Thus, we seek to verify the impact that the 'treatment effect' bullying may have on risky behaviors.

The analysis considers the victims and victimizers, separating the first ones by type of victimization. For instance, it is possible to verify the effect of bullying on risk behaviors for people who were victimized due to their color, facial appearance, body appearance, sexual orientation, religion, and region of origin. The risk activities analyzed are alcohol, substance use, sexual activity, and relation with risk peers. These behaviors were observed by factor analysis, in which variables that characterize each type of behavior were aggregated into factors, making it possible to capture the involvement of students in risk behaviors. Based on the question "Are young people involved in bullying more prone to risk?"This study hypothesizes that there is a positive effect of risky behavior for people involved in bullying experiences.

The article has three main contributions. First, by using the affective behavioral concept to analyze the decision-making process in risk engagement by students that suffer or practice bullying. Second, by analyzing the bullying effects by type of victimization, capturing different impacts of victimization on risky behavior. Finally, using the combination of factor analysis and PSM methods allows the generation of risk factors for each type of activity and enables a causal analysis.

The article is organized into four sections, in addition to this one. The first section deals with the literature review, divided into theoretical and empirical. The second section concerns the methodology, in which the data and the identification strategy are carried out. Afterward, the results are analyzed, and the last section concludes with the final remarks.

3.2 LITERATURE REVIEW

This section focuses on establishing the theoretical background used to analyze the bullying phenomenon and risky behaviors, as well as to disclose the main results verified in the empirical literature It is divided into two subsections, the first seeks to develop the theoretical aspects of affective analysis by Slovic's affect heuristic concept

[1] Heuristics are pocket rules employed by agents to facilitate decision process of complex choices (SLOVIC et al., 2007)

applied to the bullying and risky behaviors context. In the second section, contributions from the empirical literature about bullying and risky activities are addressed.

3.2.1 Theoretical Background

A positive association between bullying and risky behaviors raises a question about whether adolescents who are bullied and bully are risk-takers. A study by Flouri e Papachristou (2019) in application to the USA found that teenagers who intimidate others are more sensitive to reward (or less sensitive to punishment) than teenagers not involved in bullying. Such context instigates a more deeply analysis of risk engagement to the application of bullying circumstances.

In the decision-making process, people often rely on intuitions for risk judgment, differentiating their analysis when risks are voluntary and consequently, seen as controllable (SLOVIC, 2000a; FISCHHOFF; SLOVIC; LICHTENSTEIN, 2000). An example is that young smokers may be more susceptible to believe in the safety of short-term smoking, as they may demonstrate a degree of denial and hope to be able to stop smoking before being affected by any health damage (SLOVIC, 2000b). People may prefer to live in denial and apply this perspective to other types of risk as well, such as alcohol, sexual, and substance use, being optimistic about the consequences that exposure to these risks might have on life.

Still, thinking about consequences that they could have later in life assume a prominent role of cognition in decision making. The capacity to think through problems, knowledge of alternatives, and their implications are not the focus of risk behavior when related to bullying. On the contrary, hot affect, deep states of arousal ¬– such as fear, anger, passion – can drive people to different types of attitude, characterizing the role of affective development (GRUBER, 2001). The affective development is outstanding for bullying comprehension because it is the aspect mostly affected by this type of victimization.

The violence caused by victimizers has a source in the strong need for power and dominance. Furthermore, the negative actions to which victims of bullying are exposed repeatedly overtime can work directly, with a physical and verbal offense, or indirectly in the form of social isolation and intentional exclusion from a group (OLWEUS, 1993). In both cases, the emotional response can be drawn to provide acceptance or recognition in some group. Reliance on affect and emotion is a quicker, easier, and more efficient way to address a solution to complex and uncertain situations (FINUCANE et al., 2000).

The affective risk analysis used in this study is structured considering that individuals base their judgment not only on what they think but also on what they feel, as they can assess the risk of an activity based on the affective degree they developed with the behavior, being able to judge your risk is low if the person likes the activity and high if he or she do not like it (SLOVIC et al., 2002). People can address these conclusions by somatic markers, which can be described as alarms linked to the image of a future

outcome that can be positive or negative (FINUCANE et al., 2000; SLOVIC et al., 2007).

As victims of bullying might pursue social acceptance and aggressors might chase social recognition both can see risky behaviors as a source of integration. This is an important aspect, as young people who engage in bullying are more vulnerable to psychological problems and negative emotions (HAMMAMI et al., 2020). Also, little personal satisfaction can lead to defensive psychological attitudes towards their frustration and consequent internal aggression (SERRA-NEGRA et al., 2015). Relying on this assumption, positive markers to risk behaviors can potentially guide judgment and decision-making.

Slovic et al. (2002) label that the individual's decision-making process is divided into two types, an analytical and an experimental system. The first is logical and guided by reason, the second is holistic and affective, guided by pain and pleasure perspectives. These systems can act together on decisions making, but the salience[2] of emotional response may be more prominent for the youth. This is assumed because, as stated by Steinberg (2004), adolescents are less knowledgeable about what is risky compared to adults. Besides, the author says that the conditions of peer pressure and emotional arousal are accentuated in moments of coercion, making adolescents become more short-sighted, and highly driven by their emotions.

The affective analysis can also be structured as a model, Bracha e Brown (2012) developed a model of risky and uncertain choice with decision weights to label affective or perceived risk as endogenous. They considered two systems of reasoning, one rational and the other emotional. The rational process was responsible for the decision on action, while the emotional process formed the perception of risk, which was influenced by optimism bias[3]. The two processes interacted in the decision-making process. The main feature of this analysis is that decision-makers can have some freedom in choosing their probabilistic beliefs, being often optimistic.

Still, in Bracha e Brown (2012) the rational process is constructed with the expected utility model, in which for a given risk perception determined by their affective probability distribution, the rational process chooses an action to maximize utility, while the emotional process details the risk perspective by selecting the optimal risk perception. Furthermore, this model suggests that the failure of the expected utility model is due to systematic affective biases.

The bullying phenomenon carries a situation where these risk perceptions can be reshaped. Risk events interact with psychological, social, and cultural processes in ways that can heighten or attenuate public perception of risk (KASPERSON et al., 2000). Bullying could likely work as a social attenuator of risk. The seek for novelty and intensity in sensory experiences allied to the necessity of acceptance and recognition can influence the imagery of risky behaviors and offer a good spectrum of how youth

[2] Concept that people tend to focus their energy on things that are at the forefront of their minds (LINNEMAYR et al., 2018)

[3] Bias that people believe that their choice options carry more favorable results than it actually does.

see those behaviors as a social facilitator and as fun (ARNETT, 1994; BENTHIN et al., 2000; STEINBERG, 2004).

Finally, the importance of schools to analyze the aspects of bullying and act to avoid/eradicate such situations is severely critical to the preservation of school social capital. Once bullying phenomenon affects not only those involved but also shapes the relations of power and dominance through all the school environment, addressing communication and political enforcement to policies development is one of the main features to implement a higher level of social capital in the school.

3.2.2 Empirical Evidence

To understand the role of behavioral biases some studies can be addressed. Arnett (1994), creating a sensation seeking scale, verified that adolescents and males reported higher levels of sensation-seeking compared to adults. Besides, as demonstrated by Benthin et al. (2000), risk behaviors can usually be associated with social facilitation, having fun, physiological arousal, and physical relaxation. The sample with 411 people used in the study marked risky behaviors with feelings of goodness, happiness, and better mood, with alcohol the main social facilitator identified. Such evidence supports the idea presented by Kasperson et al. (2000), that risk-taking behaviors may serve similar psychological functions and carry similar meaning for adolescents.

O'Brien et al. (2011) analyzes an individual's preference for immediate versus delayed rewards using a delay discounting task for 100 adolescents. The authors verified that when with their friends, late adolescents aged 18-20 in the study exhibited patterns of discounting similar to those adolescents of 14-15 years old. So, youth can act immaturely in the presence of their peers even when they are older. Different aspects of involvement with bullying and risky behaviors have been documented in the literature. Considering the relevance of the subject, studies that consider bullying as one of the main aspects have increased in number, with most of the empirical cases being applied in developed countries (RIPAMONTI, 2018).

Related to the role of bullying on affective decision-making, Flouri e Papachristou (2019) identified that bullying involvement was associated with changes in decision-making in a sample of 13,888 students, even after adjustment for confounding factors. In this analysis, compared to adolescents not involved with bullying, both male and female bullies reported increased risk-taking, suggesting a higher sensitivity to rewards. This is in line with the understanding that the practice of bullying can be viewed as a component of antisocial and rule-breaking behavior patterns, such that those who engage in this activity are at increased risk of later engagement in other problem behaviors as alcohol abuse (OLWEUS, 1993).

Moreover, Forster et al. (2019) investigated associations between affective, cognitive, and behavioral engagement into three types of bullying, those are relational, physical, and perpetration. The research found out by logistic regression that affective

engagement reduces the odds of the three types of bullying behaviors. Also, males were more likely to engage in physical bullying, while females were more likely to engage in relational bullying. Serra-Negra et al. (2015) developed an interesting study for Brazilian cases using Poisson regression for 366 adolescents. They identified aggressors had low scores of self-satisfaction, so the practice of bullying could understand as a defensive psychological attitude toward their peers for the reason of their frustration.

Still, bullying might have consequences on other dimensions. For instance, in research involving 276 public high schools from Virginia (USA), Cornell et al. (2013) found that the prevalence of teasing and bullying was predictive of dropout rates for the cohort studied four years later. Also, Townsend, Flisher e King (2007) using the same method applied to Cape Town (South Africa), found that adolescents involved with bullying were at increased risk of drop out of school.

Concerning the victims' features, Malta et al. (2014), using PeNSE 2012, found that bullying was associated with male students, younger, of black color, smokers, with mental vulnerabilities, and victims of domestic violence. These findings were also found in other studies using PeNSE samples for 2009 and 2015, with a higher prevalence for bullying for those in public schools, and mothers with low levels of education (OLIVEIRA-CAMPOS et al., 2014; SILVA et al., 2019).

At the same time, Mello et al. (2018) showed that the practice of bullying is more common among students enrolled in private schools, living with both of their parents, and those mothers who have more years of schooling. A low socioeconomic status more prevalent among victims and a greater percentage of aggressors in the high socioeconomic status in these studies address a scenario where people vulnerable to bullying are also vulnerable in other aspects, such as psychological, socioeconomic, and family background, while for those who practice bullying features were more related to a demonstration of power that can already be seen in their socioeconomic status. Such a situation of vulnerability and high socioeconomic status for aggressors can deepen inequality structures, demonstrating how this phenomenon might reflect more than the school social context.

Further, it is worth mention that the report of bullying in PeNSE has increased from 5.4% in 2009 to 7.2 in 2012, varying within the margin with 7.4 in 2015. This represents a 37% increase in the prevalence of bullying between 2009 and 2015 (MELLO et al., 2018). Besides, loneliness, lack of friends, and insomnia have a higher prevalence among victims of bullying in the PeNSE 2015 sample, as well as in international experiences (SILVA et al., 2019; MALTA et al., 2019; ESCOBAR et al., 2020; TANG et al., 2020). Such aspects are recurrent in the development of internalizing problems and can contribute to potentially serious problems, as they are risk factors as self-mutilation practices, having negative consequences on psychological problems (LI et al., 2020).

It should be noted that some personal features can make young people more vulnerable to bullying. One case is that overweight and obese children are at higher risk of being bullied compared to peers with regular weight (HAMMAR; BLADH; AGNAFORS,

2020; HAMMAMI et al., 2020). Furthermore, as pointed out by Oliveira-Campos et al. (2014), with data from PeNSE 2012, the appearance of the face and body are the most cited reasons for the occurrence of bullying, though boys reported a greater proportion of bullying by race and sexual orientation.

Young people from sexual minorities may also have a greater chance of bullying and victimization by peers than young heterosexuals. A study for sexual and ethnic minorities, Jomar, Fonseca e Ramos (2020) found that the associations between bullying based on sexual orientation and feelings of loneliness/difficulty sleeping were twice as high as that found for those who were bullied by others reasons. This effect can demonstrate that some types of bullying affect young people in a particular fashion, bringing different psychological risks. Along with sexually-oriented minorities, racial/ethnic minorities are at increased risk of bullying and peer victimization also for international applications Jackman et al. (2020).

So, the factors associated with bullying involvement can make young people more prone to risky activities due to changes in affective risk perception. Studies analyzing involvement with bullying and risky behavior demonstrate a positive association between being a victim of bullying and using nicotine (MALTA et al., 2019), alcohol consumption (LEE et al., 2020), illicit drugs (ANDRADE et al., 2012; HORTA et al., 2018), and unprotected sex (MELLO et al., 2017; RAMOS et al., 2017), with features as overweight and sexual orientation contributing to a higher prevalence of bullying and risky behaviors (KUBOTA, 2015; KIM; SEO, 2020). As multivariate approaches have been widely applied in the empirical studies, this paper implements the PSM approach considering the effect of bullying on risk-controlled by the several dimensions documented in the literature to be related to a higher prevalence of bullying.

3.3 METHODOLOGY

In this section, data analysis is carried out. The data is explained, classified, inserted, and contextualized in the model. Subsequently, the seminal articles of propensity scores matching technique are used to describe the procedure on the identification strategy. After, the advantages and limitations of the study are addressed.

3.3.1 Data

To the investigation of this chapter PeNSE data contributes with a wide set of variables, divided into those of treatment, result, and control, the latter being divided into sociodemographic, school, relational, and behavior. Treatment variables relate to questions about involvement with bullying, by practice or as a victim. There are nine treatment variables, the first question is whether the young person has already suffered bullying, and does not have temporality. So, it depends on the correct understanding of the concept of bullying and may be limited by the participants' memory bias. This memory bias is a type of cognitive bias, that makes it difficult to accurately remember

events as they move away.

A fashion to address the memory bias and do not have a dependency on the understanding of bullying's concepts is to use the victimization variable. This one is presented in informal language to the students, with the use of different adjectives, hence it makes it easier for them to recognize some bullying situation. PeNSE has a formulation that considers temporality and is structured as follows "In the recent 30 days, how often did one of your schoolmates escort you, make fun of you, intimidate or tease you so much that you were hurt, bothered, upset, offended or humiliated?". The response alternatives are "never", "rarely", "sometimes", "most of the time", and "always"; being considered "most of the time", and "always" as victimized.

Another case considered is the practice of bullying by the students in the month before the research. The other six treatment binaries list the reasons given by the victims for having been victimized by bullying, they are color/race, body appearance, facial appearance, sexual orientation, religion, and region of origin. Board 5 lists all treatment variables and their characteristics.

Board 5 – Treatment variables applied in the PSM

Variable	Abbreviation	Type	Description
bullying	bullying	Binary	1= Already suffered bullying, 0= Contrary case
Victimized	victim30	Binary	1= Were victimized in the recent 30 days, 0= Contrary case.
Victimizer	victimr30	Binary	1= Practiced victimization in the recent 30 days, 0= Contrary case.
Victimization motive	color, face, body,sexorient mvrelig,region	Binaries	Six binaries related to the type of victimization suffered in the recent 30 days. Them being by color, religion, facial appearance, body appearance, sexual orientation and geographical origin.

Source: Elaborated by the author using PeNSE data.

Regarding the outcome variables, these are structured as factors that aggregate a set of variables related to the type of risky behavior. The variables used in this analysis are much the same as the one applied in Chapter 2, but there are some differences in the application of this Chapter, described briefly as follows. The types of risk behaviors used in this study are alcohol, sexual, substance use, and peers.

The alcohol factor considers the age when alcohol consumption started, the frequency which the person is drunk, the intensity at which the person is drunk, and how many times the person has been truly drunk in his life. The sexual risk factor gathers variables of sexual initiation age, number of partners, and whether the last sexual intercourse was unprotected.

The risk factor for substances relates to the age of first use of illicit drugs and the frequency of use of cigarettes, marijuana, and crack in the recent 30 days. The reason why cigarettes are considered in the substance factor occurs due to a better specification of the factorial model. Finally, the peer factor considers the number of friends who consume alcohol and the number who have used drugs.

The controls used in the model are divided into sociodemographic, school, relational, and behavioral. Sociodemographic controls bring together 21 variables, in which 18 are binary, and three are numerical. These variables relate to aspects such as sex, geographic region of origin, whether the student lives in the capital, color, age. Also, we considered aspects that seek to capture the family formation and the socioeconomic family's structure. For more information, Board 6 lists the description of each sociodemographic variable.

Board 6 – Sociodemographic controls

Variable	Abbreviation	Type	Description
Sex	sex	Binary	1= Female, 0= Male.
Region	northe,north,mid, southe,south	Binary	5 variables referring to each geographic region of Brazil. North, Northeast, Midwest, South east and south, respectively.
Capital	cap	Binary	1= Capital, 0= Interior.
Cor or Race	whi,bla,yel, bro,ind	Binary	5 variables related to the respondents' color. They are: White, Black, Yellow, Brown and indigenous, respectively.
Age	id	Numeric	Age of respondents.
Lives with father	livesfather	Binary	One lives in the same residence of the father.
Lives with mother	livesmother	Binary	One lives in the same residence as the mother.
Number household	household	Numeric	Total number of people residing in the home.
Cellphone	cell	Binary	If the respondent has a cellphone.
Internet	int	Binary	If the respondent has internet at home.
Car	car	Binary	If any family member owns a car.
Housekeeper	houkeep	Binary	If the family pays for an employee to perform housework for three days or more a week.
Works	work	Binary	If the respondent does any work paid or unpaid.
Hungry	hun30	Numeric	Likert variable from 1 to 5 representing the frequency of feeling hungry for not having enough food at home. The answers are never, rarely, sometimes, almost always and always.

Source: Elaborated by the author using PeNSE data.

Board 7 lists all other control variables. The school control variables aggregate five binary variables. This group considers whether the school where the young person studies is public, in the urban region, the maximum level of education that the young person intends to obtain, and if he has missed class more than six times in the last 30 days. The relational control variables refer mostly to the relation of young people with their families, it dealt with the frequency with which young people eat meals with parents, if any of the guardians are smokers, if they understand their problems, frequency with which they were attacked by a relative and if they had ever been raped. Finally, behavioral control variables relate to the number of hours the young person spends watching TV, if the individual feels alone, if he/she has been vaccinated against HPV, if he/she sought a health service, and what is his/her body perception, does he/she feel thin, normal or fat.

In total, there are 38 control variables used in the model. It should be emphasized that using too many covariates is not a problem because there is little risk of overfitting

Board 7 – School, relational and behavioral control variables

	Variable	Abreviation	Type	Description
School	Public	pub	Binary	1= Public School, 0= Private School.
	Urban	urb	Binary	1= Urban area, 0= Rural area.
	Maximum schooling	maxscho	Binary	1 = Aims for higher education or postgraduate studies, 0 = Aim for high school at most.
	Orientation STDs and Pregnancy	ordstpreg	Binary	1 = The school guided regarding STDs or Pregnancy. 0 = otherwise
	Missed Class	miss30	Binary	1 = In the last 30 days the student has missed classes at least six times. 0 = Missed less than six times or did not miss.
Relational	Meal with parents	agealc	Numeric	Scale from 1 to 6 reflecting the frequency with which girl with parents or guardians. 1 for every day and 6 for never.
	Parents smoke	alc_freq30	Binary	1 = At least one parent or guardian smokes, 0 = Neither parent is a smoker.
	Parents understand problems	alc_int30	Numeric	Likert variable from 1 to 5 representing the frequency that parents understand the problems. Varying never forever.
	Family aggression	drunk	Numeric	Scales from 1 to 8 with the frequency with which the young person was assaulted by a relative in the last 30 days. 1 for none and 8 for 12 or more times.
	Sexual violance	age_sex	Binary	1 = Victim of secual violance, 0 = Otherwise.
Behavior	Hours TV	sexpartn	Numeric	Scale from 1 to 10 for the number of hours watched on a typical day. Being 1 for none and 10 for more than 8 hours.
	Feels alone	preserv	Numeric	Likert variable from 1 to 5 representing the frequency that the young person feels alone. Ranging from never to 1 to always to 5.
	HPV Vaccine	drug_age	Binary	1 = Vaccinated against HPV, 0 = Otherwise.
	Health Service	cig_freq30	Binary	1 = Attended a health service in the last few for 12 months.
	Considers itself	mar_freq30	Binary	Perception of young person in relation to their body appearance. With three binaries if considered itself to be thin, normal or fat.

Source: Elaborated by the author using PeNSE data.

the data, the problem would be the use of too few, which would violate the assumption of ignorability (WONG; SCHONLAU, 2013). Besides, the PeNSE database has 102,122 students, which characterizes it as a large base. Thus, even though the data are limited by the number of observations treated in the pairing, greater security in the results can be obtained, when compared to studies that use a reduced number of data.

3.3.2 Identification Strategy

The group of people who are bullied and those who are not bullied are different in their baseline, therefore it is not clear whether a linear adjustment would be enough to produce an unbiased estimate of the treatment effect on risky behavior. As pointed out by Wong e Schonlau (2013), the bias exists because the possibility of being bullied is correlated with the individuals' characteristics. Considering that the study seeks to estimate the causal effect of victimization or bullying on the risk behavior of students, the empirical approach used follows the technique of Propensity Score Matching.

The propensity score is the conditional probability of attribution to an estimated treatment, given a vector of observed characteristics, aiming to remove the bias due to the covariates considered (ROSENBAUM; RUBIN, 1983). For the application of this technique, the hypotheses of conditional independence and common support are necessary. The first is that given the set of observed variables, the potential results must be independent of treatment. The second informs the need to have sufficiently close observations for comparability between control and treatment (MALBOUISSON; TIRYAKI, 2020).

The method of propensity score matching (PSM) is chosen so that people who are bullied and who are not bullied are comparable by creating a solid counterfactual that allows the calculation of the treatment effect. In this case, bullying works as the "treatment" and those who are not bullied are the “controls”. To estimate treatment and control, the strategy used is that of pairing by the nearest neighbor. With the same propensity scores, the only difference between treatment and control individuals will be exposed to bullying. Thus, the specification is given as follows:

$$Risk_{ij} = \beta_0 + \beta_1 treat_{ij} + X_{ij} + \epsilon_{ij} \tag{3.1}$$

In equation 3.1 risk refers to the engagement in risky behavior of individual i from school j, this risky behavior can be alcoholic, sexual, substance use, and relation with peers. The binary variable $treat$ reflects whether the student has suffered any of the types of bullying established as treatment. X is the covariate group and ϵ is the error term. The advantage of such a model is the establishment of the directionality of the effect. Commonly, episodes of drinking, drugs, tobacco use, and unprotected sex, which are interpreted as consequences, maybe a cause of it (RAMOS et al., 2017).

Among the advantages of the PSM method, the main one to be mentioned is the possibility of causal inference and removal of the bias from the observed variables. However, it cannot be controlled by unobserved variables, which might have a significant influence on the model’s result. Another limitation to the model is that the internal validity is restricted by the overlap condition and that a substantial amount of data is also required to generate the control group.

Based on a large number of covariates and the number of observations provided by PeNSE data, we proceed with the estimates. Furthermore, it is worth emphasizing that this is a cross-sectional study, then it is not possible to verify any temporal characteristics of the practice of bullying on risky behaviors. Moreover, the structure of the PeNSE self-report questionnaire may result in memory and text interpretation bias, which may be a limitation (SILVA et al., 2019). The understanding of the concepts pointed in the questionnaire and its extension implies a large number of missing answers, in this study 7,646 observations were lost when excluding missing information, resulting in a base with 94,476 observations. At this point, it is worth mentioning that the data in this chapter had a different design from chapters one and two, once some variables presented in this analysis are not considered in the other studies and the sample weights were not applied.

3.4 RESULTS

This section describes the results of the estimates made. Table 43 shows the results of the descriptive statistics of the sociodemographic control variables. The sample consists of 52.4% women, 50.4% live in capitals, 88.7% live with the mother, and 62.4% live with the father. The composition of geographic regions is slightly different

from the real one, the sample is composed mainly of young people from the north (23.4%) and northeast (35.7%). brown (46.1%) and black (12.4%) together represent 58.5% of the sample. Families are composed, on average, of 4 people at home and young people are, on average, 14 years old.

TABLE 43 – Sociodemographic descriptive statistics

Variables	Average	Std Error	Median	Minimum	Maximum	Range
Sex	0.525	0.499	1	0	1	1
North	0.234	0.423	0	0	1	1
Northeast	0.357	0.479	0	0	1	1
Center	0.138	0.345	0	0	1	1
Southeast	0.175	0.38	0	0	1	1
South	0.096	0.294	0	0	1	1
Capital	0.504	0.5	1	0	1	1
Black	0.124	0.33	0	0	1	1
White	0.332	0.471	0	0	1	1
Indígenous	0.037	0.189	0	0	1	1
Yellow	0.045	0.208	0	0	1	1
Brown	0.461	0.498	0	0	1	1
Age	14.320	1.049	14	11	19	8
Livmom	0.887	0.317	1	0	1	1
Livfat	0.624	0.484	1	0	1	1
Pdomic	4.522	1.597	4	1	10	9
Cellphone	0.875	0.331	1	0	1	1
Internet	0.774	0.418	1	0	1	1
Car	0.54	0.498	1	0	1	1
Houskee	0.118	0.323	0	0	1	1
Works	0.124	0.329	0	0	1	1
Hun30	1.365	0.779	1	1	5	4

Source: Elaborated by the author using PeNSE data.
Number of observations = 94,476

Regarding material possessions, 87.5% of young people have cellphones, 77.4% have the internet at home, 54% have a family member with a car at home, and 11.8% have a domestic employee. Besides, 12.4% of young people work and rarely claimed to be hungry at home. Table 44 shows descriptive statistics of the school, relational and behavioral control variables. Public school youth are 79% of the sample, 92.1% are from urban areas. More than half of young people (58.6%) want to go to higher education or postgraduate studies, 90% have already received guidance on sexually transmitted diseases or pregnancy at school, and 2.1% missed more than six classes in the last 30 days.

As for the relational control variables, on average, young people reported that they almost always eat with their parents, 23.8% stated that at least one of the guardians is a smoker. Besides, they stated that, on average, parents sometimes understand problems and are not attacked by family members. Also, 3.9% of young people reported having suffered sexual violence. Regarding behavior control variables, on average, young people reported spending 3.66 hours watching TV and rarely felt alone. In terms of health, 37.8% said they had been vaccinated against HPV and 56% went to a health

TABLE 44 – Control variables descriptive statistics

Controls	Variables	Average	Std Error	Median	Minimum	Maximum	Reach
School	public	0.790	0.407	1	0	1	1
	urban	0.921	0.269	1	0	1	1
	maxschoo	0.586	0.493	1	0	1	1
	ordstpreg	0.901	0.298	1	0	1	1
	miss30	0.021	0.143	0	0	1	1
Relational	refpar	2.078	1.754	1	1	6	5
	parsmo	0.238	0.426	0	0	1	1
	parent30	3.184	1.447	3	1	5	4
	agresfam30	1.342	1.120	1	1	8	7
	sexviol	0.039	0.193	0	0	1	1
Behavior	hourtv	3.669	2.704	3	0	9	9
	feelal12	2.350	1.219	2	1	5	4
	hpv	0.378	0.485	0	0	1	1
	healserv12	0.560	0.496	1	0	1	1
	thin	0.259	0.438	0	0	1	1
	normal	0.560	0.496	1	0	1	1
	fat	0.181	0.385	0	0	1	1

Source: Elaborated by the author using PeNSE data.

service in the last 12 months. Finally, 25.9% of young people have the perception that they are thin, 56% that they are normal, and 18.1% that they are fat.

Table 45 has the factor loadings of the risk factors. It can be said that it was reasonable to use a factorial model to estimate the outcome variables because the Kaiser-Meyer-Olkin coefficient was greater than 0.8. For the alcoholic risk factor, the variables with the highest load were the frequency (0.94) and intensity (0.88) of alcohol consumption. In the case of the substance factor, the nicotine use variable (0.40) obtained a lower burden compared to the age of initiation of the use of illicit drugs (0.91), frequency of marijuana (0.88), and frequency of crack (0.95). The sexual risk factor was strongly represented by the age of sexual initiation (0.90). Furthermore, the factor of friends who use alcohol (0.48) has a lower load than that of friends who use drugs (0.85).

TABLE 45 – Factor loadings - EFA result

	Alcohol	Drugs	Sex	Peers	h2	u2	comp
agealc	**0.71**	-0.13	0.08	0.02	0.51	0.49	1.10
alc_freq30	**0.94**	0.03	-0.03	-0.02	0.86	0.14	1.00
alc_int30	**0.88**	0.03	-0.02	0.00	0.79	0.21	1.00
drunk	**0.69**	0.10	0.07	0.06	0.66	0.34	1.10
cigar30	0.22	**0.40**	0.08	0.09	0.40	0.60	1.80
agedrug	0.00	**0.91**	0.03	0.00	0.84	0.16	1.00
marj30	0.02	**0.88**	0.00	0.03	0.81	0.19	1.00
crack30	-0.01	**0.95**	-0.01	-0.01	0.88	0.12	1.00
agersex	-0.02	-0.04	**0.90**	0.00	0.77	0.23	1.00
preserv	-0.01	0.03	**0.48**	0.04	0.25	0.75	1.00
pessrsex	0.03	0.07	**0.75**	-0.01	0.62	0.38	1.00
frien_alc	0.28	-0.11	0.03	**0.48**	0.40	0.60	1.70
frien_drog	-0.04	0.05	0.00	**0.85**	0.73	0.27	1.00

Source: Elaborated by the author using PeNSE data.

Table 46 shows the adequacy criteria of the factorial model. The proportion of variance explained by the factors together was 66%, with the highest proportions for the risk factor for alcohol (35%) and drugs (33%). All factor eigenvalues remained above 1, which shows that the factors are strongly explanatory.

TABLE 46 – Adequation criteria

	Alcohol	Drugs	Sexual	Peers
SS factors	2.94	2.79	1.70	1.09
Proportion of variance	0.23	0.21	0.13	0.08
Accumulated variance	0.23	0.44	0.57	0.66
Ratio explained	0.35	0.33	0.20	0.13
Accumulated Ratio	0.35	0.67	0.87	1.00

Source: Elaborated by the author using PeNSE data.

After building the factor analysis model, we proceed with the PSM estimates. All estimates and balances are present in Tables 90 - 98 from Appendix C. Figure 10 shows the distribution of propensity scores for treated and control before and after pairing for the bullying variable. It presents a reasonable overlap of the scores, as well as an improvement in the perceived means. Figure 11 is the overlapping of the pairing considering the victimized treatment variable, in which a better post-matching overlap was obtained.

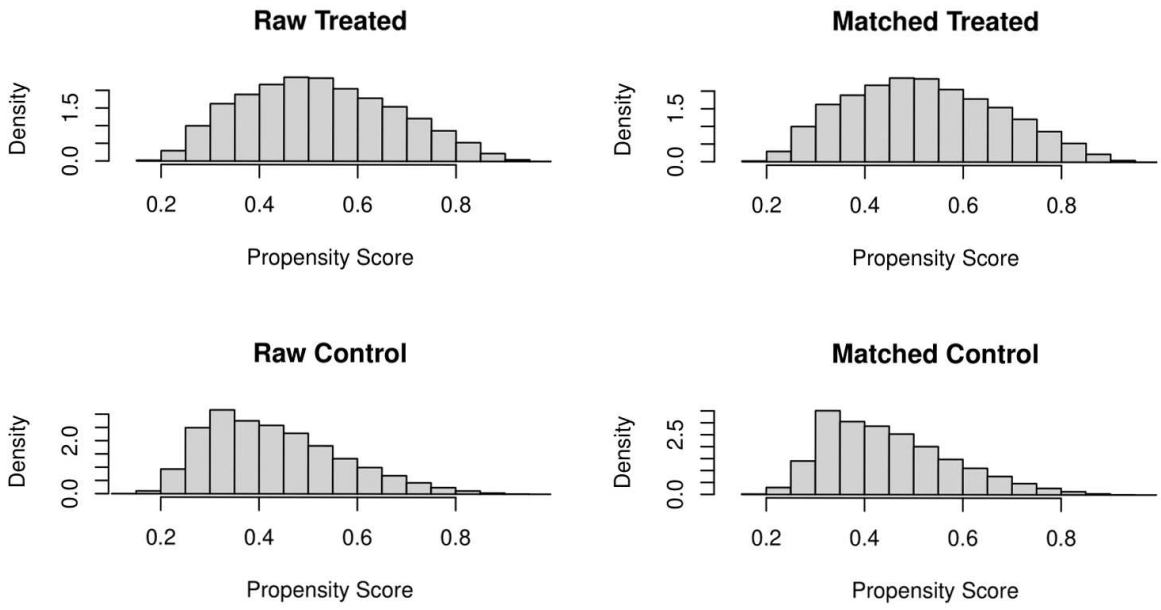

FIGURE 10 – Bullying PSM

Finally, Figure 12 shows the distribution of propensity scores for the victimizer variable, in which it is possible to identify a good overlap of the data. The figures related to the quality of the PSMs overlapping by type of victimization are present in Figures 13 - 18, Appendix C. Table 47 presents the results of difference in means of the treatment effect of bullying, victimized and victimizer. In most of the estimates, having some

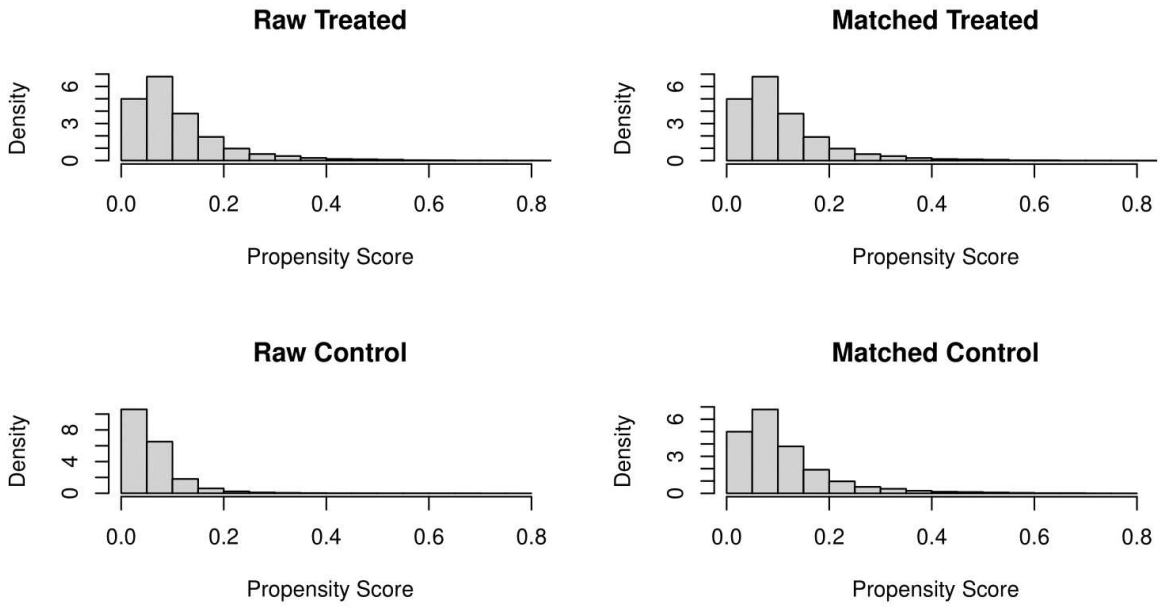

FIGURE 11 – Victimized PSM

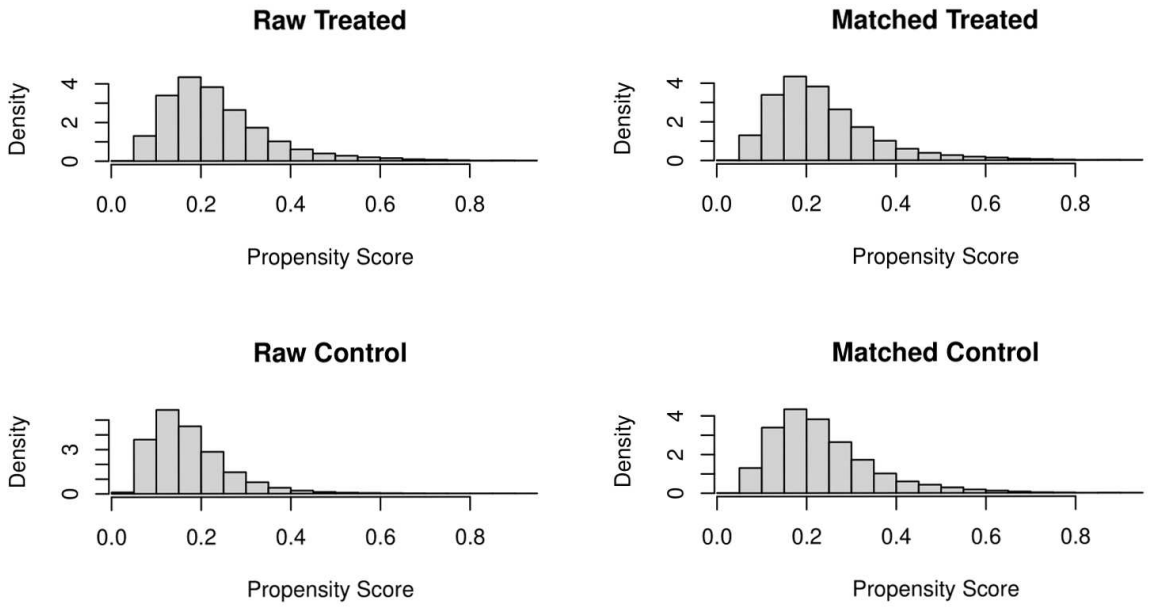

FIGURE 12 – Victimizer PSM

kind of involvement with bullying had a positive effect on risky behavior. In the case of the bullying treatment variable, all differences were significant at the level of 0.01, even including the controls, with the highest magnitudes being found for alcoholic risk behavior (0.1042) and relation with risk peers (0.1268).

The victim treatment variable registered a high significance in the first estimate, which considers only sociodemographic controls. However, adding new controls, some differences lost significance. In this case, the only behavior that remained significant in

TABLE 47 – Difference in means of risk behaviors' treatment effect

Risk Factors		Bullying		
Alcoholic	0.1234***	0.1233***	0.1075***	0.1042***
Sexual	0.0436***	0.0405***	0.0315***	0.0247***
Drugs	0.0722***	0.0722***	0.0632***	0.0631***
Peers	0.1428***	0.1426***	0.1305***	0.1268***
n	89,544	89,544	89,544	89,544
Risk Factors		Victimized		
Alcoholic	0.1407***	0.1099***	0.0282	-0.0107
Sexual	0.0540***	0.0133	-0.0448***	-0.0547***
Drugs	0.0861***	0.0826***	-0.0049	-0.0271
Peers	0.2011***	0.1818***	0.0932***	0.0655***
n	12,452	12,452	12,452	12,452
Risk Factors		Victimizer		
Alcoholic	0.5113***	0.4976***	0.4320***	0.3948***
Sexual	0.2999***	0.2869***	0.2400***	0.2242***
Drugs	0.3631***	0.3595***	0.3234***	0.3038***
Peers	0.4794***	0.4737***	0.4102***	0.3901***
n	34,964	34,964	34,964	34,964
Controls				
Demographic	Yes	Yes	Yes	Yes
School	No	Yes	Yes	Yes
Relational	No	No	Yes	Yes
Behavior	No	No	No	Yes

Source: Elaborated by the author using PeNSE data.
Note: *** p<0.01. ** p<0.05. * p<0.1

all differences was risk peers, which was positive, although it lost its magnitude when new controls were inserted.

The mean differences for the treatment effect of the victimizer variable shows the high significance and strong magnitudes in all cases. The biggest differences are given by alcoholic risk behavior (0.3948) and relation with risk peers (0.3901). However, the magnitudes of sexual risk behaviors (0.2242) and substance use (0.3038) were also high in the case of victimizers. This question demonstrates that the victimizers may be at risk more sharply than the victims of the practice of bullying, which manifests a specificity of the domination relations that have effects in other branches.

Table 48 presents the results of the difference in means of the treatment effect of the different types of victimization. Some differences can be pointed out for each type of victimization; however, for the most part, the differences are positive. In the case of victimization by color or race, the risk behaviors that remained significant in all estimates were alcohol and peer relations. When considering all controls, the impact of victimization by color on alcoholic risk behavior and concerning to peers was 0.0664 and 0.1150, respectively.

Victimization due to body appearance had a positive impact on the relation with risk peers in all controls. On the other hand, there was a negative impact on sexual risk

TABLE 48 – Difference in means for treatment effect on types of victimization on risk behaviors

Risk Factor	Color			
Alcoholic	0.1463***	0.1580***	0.0891***	0.0664**
Sexual	0.0585*	0.0599*	0.0270	-0.0075
Substancies	0.0724**	0.0369	0.0319	-0.0383
Peers	0.2031***	0.1700***	0.1539***	0.1150***
n	4,182	4,182	4,182	4,182
Risk Factor	Body			
Alcoholic	0.0668***	0.0883***	0.0367**	-0.0137
Sexual	-0.0062	0.0014	-0.0394***	-0.0610***
Substancies	0.0281*	0.0375**	-0.0255	-0.0421**
Peers	0.1128***	0.1336***	0.0812***	0.0431***
n	12,918	12,918	12,918	12,918
Risk Factor	Face			
Alcoholic	0.0688***	0.0556***	0.0143	-0.0187
Sexual	-0.0019	-0.0018	-0.0429**	-0.0467**
Substancies	0.0391*	0.0390*	-0.0017	-0.0381
Peers	0.0939***	0.1027***	0.0425***	0.0256
n	8,316	8,316	8,316	8,316
Risk Factor	Sexual Orientation			
Alcoholic	0.6212***	0.5720***	0.4344***	0.3639***
Sexual	0.4964***	0.4571***	0.2996***	0.3434***
Substancies	0.4838***	0.4986***	0.3072***	0.1992**
Peers	0.5317***	0.4930***	0.3558***	0.2813***
n	1,602	1,602	1,602	1,602
Risk Factor	Religion			
Alcoholic	-0.1596***	-0.1727***	-0.1511***	-0.1482***
Sexual	-0.0587	-0.0576	-0.0684*	-0.0788**
Substancies	0.0109	0.0343	-0.0521	0.0032
Peers	0.0443	0.0629*	0.0438	0.0551
n	2,766	2,766	2,766	2,766
Risk Factor	Region			
Alcoholic	0.0811	0.0958	0.0572	-0.0198
Sexual	-0.0018	0.1190**	0.0703	-0.0142
Substancies	0.0875	0.1452*	0.0888	0.1078
Peers	0.1308**	0.1497**	0.0494	0.0184
n	1,094	1,094	1,094	1,094
Controls				
Demographic	Yes	Yes	Yes	Yes
School	No	Yes	Yes	Yes
Relational	No	No	Yes	Yes
Behavior	No	No	No	Yes

Source: Elaborated by the author using PeNSE 2015.
Obs.: *** p<0.01. ** p<0.05. * p<0.1

behavior (-0.061) and substance use (-0.0421) when considering all controls, but it was not significant when considering sociodemographic and/or school controls. In the case of victimization due to facial appearance, the only risky behavior that was significant, considering all controls, was sexual, which obtained a negative result (-0.0467). The relation with peers proved to be significant until controlled by the relational controls, but lost significance when adding the behavioral ones.

Bullying motivated by sexual orientation proved to be a differential in estimates, as it obtained positive and significant results for all differences. The most significant behaviors were alcohol (0.3639) and sexual (0.3434). It is worth noting that such magnitudes manifest a specificity of this type of victimization, which may reflect deeper problems intrinsic to the consequences of victimization by sexual orientation.

Differences due to religion showed mostly negative results when significant, the only exception was the relation of peers when controlled for sociodemographic and school variables, the behavior of alcoholic risk was the only case with significance in all controls. Finally, victimization by region of origin proved to be without significant results when considered all controls.

3.5 CONCLUSION

Risk behavior engagement is a relevant subject because it affects young people's safety and health with consequences in other stages of life. This study sought to analyze the impact of bullying on risk behaviors considering the affective analysis, which is based on the affect heuristic. The results showed that there is a recurrence in the positive impact of bullying practices on risk behaviors, with stronger magnitudes for victimizers and victims based on sexual orientation.

The results obtained by the differences in risk behavior of young victimizers are in line with the study by Flouri e Papachristou (2019), which identifies that young people who practice victimization are less sensitive to losses, with reward-seeking behavior. Besides, in almost all estimates, alcohol and relationships with risk peers were significant and positive, this is because an affective analysis of alcohol risk may result in establishing relationships and a network of friends, in addition to being the most common type of consumption socially accepted. As found by Escobar et al. (2020) high alcohol consumption was associated with few friends and moderate consumption associated with more relationships, so that depending on how young people approach such a risk, this practice can be understood as beneficial.

Victimization based on students' sexual orientation had strong magnitudes and these results corroborate previous studies that demonstrated that sexual minorities are at greater risk of being bullied (JACKMAN et al., 2020) and that these people can be affected in a particular way because this practice is related to a deeper form of psychological risk. According to Jomar, Fonseca e Ramos (2020), one can draw attention to the fact that the more parents understand their children's problems and

concerns and the more close friends they have, the lower the prevalence of guidance-based bullying sexual. Such differences can occur due to the protective network of these young people, a group of friends or family members who monitor and care about young people can be a key factor to avoid such practices, however, sexual minorities may not receive the same type of protection from these

Another particular case was that victimization by religion resulted in a negative impact of alcoholic risk behavior, this can be understood because young religious people have a well-established network that rejects such practices, and therefore serve as protection. Besides, in some specific types of victimization (body, face, and sexual) they harmed the practice of bullying when adding all controls, one can understand this result considering that the last group of controls considers body perception variables that are relevant for the results of such an effect. Thus, once controlled, it was possible to verify a negative impact, which may occur due to the marginalization of victimized people due to their appearance, making the establishment of emotional relationships with other people more complex.

Considering the results presented, we interpret that the affective risk analysis of young people exposed to victimization or its practice adds a different perception of risk, which can make the marginal benefit of involvement in risky behavior greater than its marginal cost. Also, some characteristics are relevant to explain the phenomenon, since the establishment of protective networks can bring young people closer to risky practices, see religion and the relationship with peers who consume alcohol or use drugs.

This study has potential implications in terms of public policies, as establishing bullying as a direction for risk engagement makes policies that seek to suppress such practices also have implications for the use of alcohol, substances, and involvement with peers risk. In this sense, psychology can be useful for structuring policies that aim to reduce such behavioral and health problems, to inform the use and design of traditional political tools for behavior change (MADRIAN, 2014).

Related to policy application, it is important to address how health professionals can establish a better school environment and provide better health practices among young people, to make possible a commitment and action concerning violence, in which situations of bullying school examples are increasingly present and frequent (SILVA et al., 2019). Campaigns aimed at reducing the practice of victimization can be beneficial and schools can establish practices that seek to generate systematic falls in the occurrences of bullying among schools.

CONCLUSION

This dissertation aimed to analyze aspects likely to become a source of concern to students. We explored characteristics related to schooling expectations, such as socioeconomic, relational/behavioral, and risk aspects; addressing interpretations based on the theoretical approach underlined. The three chapters used PeNSE data and analyzed related features in the papers. That is, the first chapter had unclear results related to risk variables' role on schooling expectation. The second chapter applied a different methodology to improve risk behavior analysis. The third chapter investigated the role of bullying, which was demonstrated to be positively correlated with risk behaviors and schooling expectations in the second chapter.

The first chapter had an exploratory focus and demonstrated that features related to a vulnerable condition are likely to play an important role in students' expectations. At the same time, relational and risk problems are related to lower odds of having high expectations. Also, we raised a discussion on how to apply policies regarding schooling expectations. The central analysis is that policies need to establish an environment where students have support and expectations can be achieved. The main contribution of this chapter relies on the schooling expectation analysis related to a wide set of variables, which supported the interpretation that students' prospects may be highly influenced by the environment and circumstances in which they live.

The second essay investigated the relation between schooling expectation and risk behavior by another approach. We relied on behavioral economic theory to analyze students' motives to engage in risky behaviors and analyze how they can be related to schooling expectations. Concepts as present bias, myopia, and visceral effect had a relevant role in the interpretation of results. Moreover, the analysis implemented draws attention to the self-control problem. Young people may have misleading perspectives of future behavior and do not recognize biases/inconsistencies in their behavior. As a result, they act naively and might systematically predict their behavior incorrectly. On the other hand, there may be young people that consider self-control problems and try to develop mechanisms to prevent future bad behavior; acting as sophisticates.

The analysis implemented supports that risky behaviors may be negative to schooling expectations. These risky behaviors were identified with constructs, aggregating the observed variables from each risk type, e.g. alcohol, drugs, and sex relations. Our results identified that risky sexual behavior was consistently negatively correlated to schooling expectations, while the other constructs presented a variation in the results. Thus, the risky sexual behavior was interpreted as sophisticated and the others as naively perceived. Accordingly, we interpreted that policies should address self-control problems management and mechanisms to prevent students from accessing such risky behaviors.

Lastly, the third essay focused on the relationship between bullying and risky behaviors. The propensity score matching approach was used to verify how the bullying and victimization process might affect risky activities. Affective risk analysis was used as the theoretical background, with concepts as affect heuristic, salience, and emotional markers being used to support the interpretations. The results indicated that bullying involvement positively affects risky behaviors in a major part of significant cases. Additionally, the magnitudes were higher for students who practice bullying and those bullied due to their sexual orientation. As a result, policies aiming to diminish bullying are critical to students' well-being and may impact risky behaviors.

The dissertation contributed to the decision-making process subject focusing on students from ninth grade aged 14-15 years old. A similar analysis of all chapters was that environments with higher acceptance were related to lower evidence of risk behaviors, low schooling expectation, and bullying involvement. Moreover, enforcement of the implementation of higher levels of social capital in schools may work as a protective factor for students. Besides, human development is an aspect of critical analysis and their ´perceptions must be taken into account. To conclude, once more we address that incentives matter and they are a key point to implement a change on youths' perceptions.

REFERENCES

AGRESTI, A. *An Introduction to Categorical Data Analysis*. 3rd. ed. [S.l.]: John Wiley & Sons, 2019. (Wiley Series in Probability and Statistics). ISBN 9781119405276.

AINSLIE, G.; HASLAM, N. Self-control. In: *Choice Over Time*. [S.l.]: Russell Sage Foundation, 1992.

ALMEIDA, A. T. C. d.; ARAÚJO-JÚNIOR, I. T. d. Efeitos da exposição aos fatores de risco comportamentais à saúde sobre o atraso escolar no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 70, n. 2, p. 129–169, 2016.

ANDRADE, S. S. C. d. A. et al. Relação entre violência física, consumo de álcool e outras drogas e bullying entre adolescentes escolares brasileiros. *Cadernos de Saúde Pública*, SciELO Public Health, v. 28, p. 1725–1736, 2012.

ARNETT, J. Sensation seeking: A new conceptualization and a new scale. *Personality and individual differences*, Elsevier, v. 16, n. 2, p. 289–296, 1994.

ATTANASIO, O. P.; KAUFMANN, K. M. Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, Elsevier, v. 109, p. 203–216, 2014.

BECKER, G. S. *Human Capital : a theoretical and empirical analysis, with special reference to education*. London: The University of Chicago Press, 1993.

BECKER, G. S.; MURPHY, K. M. A theory of rational addiction. *Journal of political Economy*, The University of Chicago Press, v. 96, n. 4, p. 675–700, 1988.

BENTHIN, A. et al. Adolescent health-threatening and health-enhancing behaviors: A study of word association and imagery. In: *The Perception of Risk*. [S.l.]: Taylor & Francis, 2000. p. 316–326.

BICKEL, W. K.; MOODY, L.; HIGGINS, S. T. Some current dimensions of the behavioral economics of health-related behavior change. *Preventive medicine*, Elsevier, v. 92, p. 16–23, 2016.

BOLLEN, K. A.; NOBLE, M. D. Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 108, n. Supplement 3, p. 15639–15646, 2011.

BRACHA, A.; BROWN, D. J. Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, Elsevier, v. 75, n. 1, p. 67–80, 2012.

BUCHANAN, E. M. *Advanced Statistics*. [S.l.]: CC-By Attribution 4.0 International, 2020. DOI 10.17605/OSF.IO/DNUYV.

CARD, D.; LEMIEUX, T. Dropout and enrollment trends in the postwar period: What went wrong in the 1970s? In: *Risky behavior among youths: An economic analysis*. [S.l.]: University of Chicago Press, 2001. p. 439–482.

CHIAPA, C.; PRINA, S.; PARKER, A. The effects of financial inclusion on children's schooling, and parental aspirations and expectations. *Journal of International Development*, Wiley Online Library, v. 28, n. 5, p. 683–696, 2016.

COLEMAN, J. S. Social capital in the creation of human capital. *American journal of sociology*, University of Chicago Press, v. 94, p. S95–S120, 1988.

COMBER, B. Literacy, poverty and schooling: what matters in young people's education? *Literacy*, Wiley Online Library, v. 48, n. 3, p. 115–123, 2014.

CORNELL, D. et al. Perceived prevalence of teasing and bullying predicts high school dropout rates. *Journal of educational psychology*, American Psychological Association, v. 105, n. 1, p. 138, 2013.

COWAN, B. W. Forward-thinking teens: The effects of college costs on adolescent risky behavior. *Economics of education review*, Elsevier, v. 30, n. 5, p. 813–825, 2011.

DHAMI, M. K.; MANDEL, D. R. Crime as risk taking. *Psychology, crime & law*, Taylor & Francis, v. 18, n. 4, p. 389–403, 2012.

DIAZ, A. et al. Effect of child abuse and neglect on risk behaviors in inner-city minority female adolescents and young adults. *Child abuse & neglect*, Elsevier, v. 101, p. 104347, 2020.

DO, H. N. et al. Patterns of risky sexual behaviors and associated factors among youths and adolescents in vietnam. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 17, n. 6, p. 1903, 2020.

DO, Y. K.; WONG, K. Y. Awareness and acceptability of human papillomavirus vaccine: an application of the instrumental variables bivariate probit model. *BMC public health*, Springer, v. 12, n. 1, p. 1–8, 2012.

ELDER, G. H. The life course as developmental theory. *Child development*, Wiley Online Library, v. 69, n. 1, p. 1–12, 1998.

EPSKAMP, S. *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output*. [S.l.], 2019. R package version 1.1.2. Disponível em: <https://CRAN.R-project.org/package=semPlot>.

ESCOBAR, D. F. S. S. et al. Assessing the mental health of brazilian students involved in risky behaviors. *International Journal of Environmental Research and Public Health*, Multidisciplinary Digital Publishing Institute, v. 17, n. 10, p. 3647, 2020.

FAVARA, M. Do dreams come true? aspirations and educational attainments of ethiopian boys and girls. *Journal of African Economies*, Oxford University Press, v. 26, n. 5, p. 561–583, 2017.

FINUCANE, M. L. et al. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, Wiley Online Library, v. 13, n. 1, p. 1–17, 2000.

FISCHER, C. *Read this paper even later: Procrastination with time-inconsistent preferences*. [S.l.], 1999. Discussion Paper Series 99-20. Research in Agricultural and Applied Economics. Disponível em: <http://ageconsearch.umn.edu/record/10725>.

FISCHHOFF, B.; SLOVIC, P.; LICHTENSTEIN, S. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, American Psychological Association, v. 3, n. 4, p. 552, 1977.

FISCHHOFF, B.; SLOVIC, P.; LICHTENSTEIN, S. Weighing the risks: Risks: Benefits which risks are acceptable? *Environment: Science and Policy for Sustainable Development*, Taylor & Francis, v. 21, n. 4, p. 17–38, 1979.

FISCHHOFF, B.; SLOVIC, P.; LICHTENSTEIN, S. Weighing the risks: Which risks are acceptable? In: *The perception of risk*. [S.l.]: Earthscane, Routledge, 2000. p. 121–136.

FLOURI, E.; PAPACHRISTOU, E. Peer problems, bullying involvement, and affective decision-making in adolescence. *British journal of developmental psychology*, Wiley Online Library, v. 37, n. 4, p. 466–485, 2019.

FORSTER, M. et al. Associations between dimensions of school engagement and bullying victimization and perpetration among middle school students. *School mental health*, Springer, p. 1–12, 2019.

FRANÇA, M. T. A.; FRIO, G. S. Factors associated with family, school and behavioral characteristics on sexual initiation: A gender analysis for brazilian adolescents. *PloS one*, Public Library of Science San Francisco, CA USA, v. 13, n. 12, p. e0208542, 2018.

FREDERICK, S.; LOEWENSTEIN, G.; O'DONOGHUE, T. Time discounting and time preference: A critical review. In: *Advances in Behavioral Economics Edited*. [S.l.]: Princeton University Press, 2004. p. 162–222.

FULLER, W. C.; MANSKI, C. F.; WISE, D. A. New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources*, JSTOR, p. 477–498, 1982.

GIUSTINELLI, P.; MANSKI, C. F. Survey measures of family decision processes for econometric analysis of schooling decisions. *Economic Inquiry*, Wiley Online Library, v. 56, n. 1, p. 81–99, 2018.

GLAESER, E. L. Why does schooling generate economic growth? *Economics Letters*, Elsevier, v. 44, n. 3, p. 333–337, 1994.

GOULET, M. et al. Longitudinal association between risk profiles, school dropout risk, and substance abuse in adolescence. In: SPRINGER. *Child & Youth Care Forum*. [S.l.], 2020. v. 49, n. 5, p. 687–706.

GRUBER, J. *Risky Behavior Among Youths: An Economic Analysis*. 1. ed. [S.l.]: University Of Chicago Press, 2001. ISBN 0226310132,9780226310138,9780226309972.

HAIR, J. F. et al. *Análise Multivariada de Dados*. 6th. ed. [S.l.]: Bookman, 2009.

HAIR, J. F. et al. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd. ed. [S.l.]: Sage, 2017.

HAMMAMI, N. et al. Exploring gender differences in the longitudinal association between bullying and risk behaviours with body mass index among compass youth in canada. *Preventive medicine*, Elsevier, v. 139, p. 106188, 2020.

HAMMAR, E.; BLADH, M.; AGNAFORS, S. Mental health and experience of being bullied in 12-year-old children with overweight and obesity. *Acta Paediatrica*, Wiley Online Library, v. 109, n. 7, p. 1450–1457, 2020.

HARRELL, J. a. F. E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2. ed. [S.l.]: Springer International Publishing, 2015. (Springer Series in Statistics).

HARRISON, A. et al. Cigarette smoking, mental health, and other substance use among court-involved youth. *Substance use & misuse*, Taylor & Francis, v. 55, n. 4, p. 572–581, 2020.

HEARD, H. E. Fathers, mothers, and family structure: Family trajectories, parent gender, and adolescent schooling. *Journal of Marriage and Family*, Wiley Online Library, v. 69, n. 2, p. 435–450, 2007.

HEINECK, G.; SCHWARZE, J. Substance use and earnings: the case of smokers in germany. *Available at SSRN 392061*, 2003. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=392061>.

HERRNSTEIN, R. J.; PRELEC, D. A theory of addiction. Russell Sage Foundation, 1992.

HOOPER, D.; COUGHLAN, J.; MULLEN, M. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, v. 6, n. 1, p. 53–60, 2008. Disponível em: <DOI:10.21427/D7CF7R>.

HORTA, R. L. et al. Prevalence and factors associated with illicit drug use throughout life: National school health survey 2015. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 21, p. e180007, 2018.

IBGE. *Pesquisa Nacional de Saúde do Escolar: 2015. Instituto Brasileiro de Geografia e Estatítica - Coordenação de População e Indicadores Sociais*. Rio de Janeiro, 2016. 132 p. ISBN 978-85-240-4387-1.

ISAKSSON, J. et al. Risk factors associated with alcohol use in early adolescence among american inner-city youth: A longitudinal study. *Substance use & misuse*, Taylor & Francis, v. 55, n. 3, p. 358–366, 2020.

JACKMAN, K. et al. Bullying and peer victimization of minority youth: intersections of sexual identity and race/ethnicity. *Journal of school health*, Wiley Online Library, v. 90, n. 5, p. 368–377, 2020.

JOMAR, R. T.; FONSECA, V. A. de O.; RAMOS, D. de O. Effects of sexual orientation-based bullying on feelings of loneliness and sleeping difficulty among brazilian middle school students. *Jornal de pediatria*, Elsevier, 2020.

KASPERSON, R. E. et al. The social amplification of risk: A conceptual framework. In: *The perception of risk*. [S.l.]: Earthscane, Routledge, 2000. p. 232–245.

KIM, J.-S.; SEO, Y. Victimization as a mediator in the relationship between sexual orientation and adolescent alcohol use. *Archives of psychiatric nursing*, Elsevier, v. 34, n. 1, p. 27–34, 2020.

KUBOTA, L. C. Discrimination against the obese and very thin students in brazilian schools. Instituto de Pesquisa Econômica Aplicada (Ipea), 2015.

LANGER, E. J. The illusion of control. *Journal of personality and social psychology*, American Psychological Association, v. 32, n. 2, p. 311, 1975.

LAUGLO, J.; LIU, F. The reverse gender gap in adolescents' expectation of higher education: Analysis of 50 education systems. *Comparative Education Review*, University of Chicago Press Chicago, IL, v. 63, n. 1, p. 28–57, 2019.

LEE, J. et al. Exploring sex differences in the association between bullying involvement and alcohol and marijuana use among us adolescents in 6th to 10th grade. *Substance use & misuse*, Taylor & Francis, v. 55, n. 8, p. 1203–1213, 2020.

LI, X. et al. Research on the relationships between psychological problems and school bullying and non-suicidal self-injury among rural primary and middle school students in developing areas of china. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 17, n. 10, p. 3371, 2020.

LINNEMAYR, S. et al. Using behavioral economics to promote hiv prevention for key populations. *Journal of AIDS & clinical research*, NIH Public Access, v. 9, n. 11, 2018.

LOEWENSTEIN, G. Out of control: Visceral influences on behavior. In: *Advances in Behavioral Economics Edited*. [S.l.]: Princeton University Press, 2004. p. 689–724.

LOEWENSTEIN, G.; PRELEC, D. Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, MIT Press, v. 107, n. 2, p. 573–597, 1992.

LUMLEY, T. Analysis of complex survey samples. *Journal of Statistical Software*, v. 9, n. 1, p. 1–19, 2004. R package verson 2.2.

LUMLEY, T. *Complex Surveys: A Guide to Analysis Using R: A Guide to Analysis Using R*. [S.l.]: John Wiley and Sons, 2010.

MADRIAN, B. C. Applying insights from behavioral economics to policy design. *Annu. Rev. Econ.*, Annual Reviews, v. 6, n. 1, p. 663–688, 2014.

MALBOUISSON, C. S.; TIRYAKI, G. F. *Econometria na Prática*. Rio de Janeiro: Alta Books, 2020.

MALTA, D. C. et al. Prevalence of bullying and associated factors among brazilian schoolchildren in 2015. *Ciencia & saude coletiva*, SciELO Public Health, v. 24, p. 1359–1368, 2019.

MALTA, D. C. et al. Bullying e fatores associados em adolescentes brasileiros: análise da pesquisa nacional de saúde do escolar (pense 2012). *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 17, p. 131–145, 2014.

MANSKI, C. F. Adolescent econometricians: How do youth infer the returns to schooling? In: *Studies of supply and demand in higher education*. [S.l.]: University of Chicago Press, 1993. p. 43–60.

MANSKI, C. F. Measuring expectations. *Econometrica*, Wiley Online Library, v. 72, n. 5, p. 1329–1376, 2004.

MELLO, F. C. M. et al. Evolution of the report of suffering bullying among brazilian schoolchildren: National scholl health survey-2009 to 2015. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 21, p. e180015, 2018.

MELLO, F. C. M. et al. A prática de bullying entre escolares brasileiros e fatores associados, pesquisa nacional de saúde do escolar 2015. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 22, p. 2939–2948, 2017.

MELLO, Z. R. Gender variation in developmental trajectories of educational and occupational expectations and attainment from adolescence to adulthood. *Developmental psychology*, American Psychological Association, v. 44, n. 4, p. 1069, 2008.

MIHALEC-ADKINS, B. P.; COOLEY, M. E. Examining individual-level academic risk and protective factors for foster youth: School engagement, behaviors, self-esteem, and social skills. *Child & Family Social Work*, Wiley Online Library, v. 25, n. 2, p. 256–266, 2020.

NELSON, J. P. What is learned from longitudinal studies of advertising and youth drinking and smoking? a critical assessment. *International Journal of Environmental Research and Public Health*, Molecular Diversity Preservation International, v. 7, n. 3, p. 870–926, 2010.

OBERSKI, D. lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, v. 57, n. 1, p. 1–27, 2014. Disponível em: <http://www.jstatsoft.org/v57/i01/>.

O'BRIEN, L. et al. Adolescents prefer more immediate rewards when in the presence of their peers. *Journal of Research on adolescence*, Wiley Online Library, v. 21, n. 4, p. 747–753, 2011.

O'DONOGHUE, T.; RABIN, M. Doing it now or later. *American economic review*, v. 89, n. 1, p. 103–124, 1999.

OLIVEIRA-CAMPOS, M. et al. Comportamento sexual em adolescentes brasileiros, pesquisa nacional de saúde do escolar (pense 2012). *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 17, p. 116–130, 2014.

OLWEUS, D. *Bullying at School: What We Know and What We Can Do*. 1st. ed. [S.l.]: Wiley-Blackwell, 1993.

O'DONOGHUE, T.; RABIN, M. Risky behavior among youths: Some issues from behavioral economics. In: *Risky behavior among youths: An economic analysis*. [S.l.]: University of Chicago Press, 2001. p. 29–68.

O'DONOGHUE, T.; RABIN, M. Addiction and present-biased preferences. UC Berkeley: Department of Economics., 2002. Disponível em: <https://escholarship.org/uc/item/3v86x53j>.

RABIN, M. Psychology and economics. *Journal of economic literature*, JSTOR, v. 36, n. 1, p. 11–46, 1998.

RAMOS, D. d. O. et al. The role of city income inequality, sex ratio and youth mortality rates in the effect of violent victimization on health-risk behaviors in brazilian adolescents. *Social Science & Medicine*, Elsevier, v. 181, p. 17–23, 2017.

REVELLE, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois, 2019. R package version 1.9.12. Disponível em: <https://CRAN.R-project.org/package=psych>.

RIPAMONTI, E. Risk factors for dropping out of high school: A review of contemporary, international empirical research. *Adolescent Research Review*, Springer, v. 3, n. 3, p. 321–338, 2018.

ROGERS, C. J. et al. The role of perceived discrimination in substance use trajectories in hispanic young adults: A longitudinal cohort study from high school through emerging adulthood. *Addictive behaviors*, Elsevier, v. 103, p. 106253, 2020.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Oxford University Press, v. 70, n. 1, p. 41–55, 1983.

ROSSEEL, Y. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, v. 48, n. 2, p. 1–36, 2012. Disponível em: <http://www.jstatsoft.org/v48/i02/>.

SCHWARTZMAN, S.; BROCK, C. Os desafios da educação no brasil. *Os desafios da educação no Brasil. Rio de Janeiro: Nova Fronteira*, p. 9–51, 2005.

SEEFELDT, C.; GALPER, A.; DENTON, K. Head start children's conceptions of and expectations for their future schooling. *Early Childhood Research Quarterly*, Elsevier, v. 12, n. 4, p. 387–406, 1997.

SERRA-NEGRA, J. M. et al. Verbal school bullying and life satisfaction among brazilian adolescents: Profiles of the aggressor and the victim. *Comprehensive Psychiatry*, Elsevier, v. 57, p. 132–139, 2015.

SILVA, J. L. d. et al. Prevalência da prática de bullying referida por estudantes brasileiros: dados da pesquisa nacional de saúde do escolar, 2015. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 28, p. e2018178, 2019.

SINGH, S. et al. E-cigarettes and youth: Patterns of use, potential harms, and recommendations. *Preventive Medicine*, Elsevier, v. 133, p. 106009, 2020.

SLOVIC, P. Perception of risk. In: *The perception of risk*. [S.l.]: Earthscane, Routledge, 2000. p. 220–231.

SLOVIC, P. What does it mean to know a cumulative risk? adolescents' perceptions of short-term and long-term consequences of smoking. *Journal of behavioral decision making*, Wiley Online Library, v. 13, n. 2, p. 259–266, 2000.

SLOVIC, P. et al. Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*, Elsevier, v. 31, n. 4, p. 329–342, 2002.

SLOVIC, P. et al. The affect heuristic. *European journal of operational research*, Elsevier, v. 177, n. 3, p. 1333–1352, 2007.

SMITH-GREENAWAY, E.; YEATMAN, S. Unrealized educational expectations and mental health: Evidence from a low-income country. *Social Forces*, Oxford University Press, v. 98, n. 3, p. 1112–1142, 2020.

SMYTH, E. Shaping educational expectations: The perspectives of 13-year-olds and their parents. *Educational Review*, Taylor & Francis, v. 72, n. 2, p. 173–195, 2020.

STEINBERG, L. Risk taking in adolescence: what changes, and why? *Annals of the New York Academy of Sciences*, Wiley Online Library, v. 1021, n. 1, p. 51–58, 2004.

STROTZ, R. H. Myopia and inconsistency in dynamic utility maximization. *The review of economic studies*, JSTOR, v. 23, n. 3, p. 165–180, 1955.

SUTHERLAND, I.; SHEPHERD, J. P. Social dimensions of adolescent substance use. *Addiction*, Wiley Online Library, v. 96, n. 3, p. 445–458, 2001.

TANG, W. et al. Impact of parental absence on insomnia and nightmares in chinese left-behind adolescents: A structural equation modeling analysis. *Children and Youth Services Review*, Elsevier, v. 114, p. 105076, 2020.

TOWNSEND, L.; FLISHER, A. J.; KING, G. A systematic review of the relationship between high school dropout and substance use. *Clinical child and family psychology review*, Springer, v. 10, n. 4, p. 295–317, 2007.

VOGT, W. P. *Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences*. London: SAGE, 2005.

WALTERS, G. D. Prosocial peers as risk, protective, and promotive factors for the prevention of delinquency and drug use. *Journal of youth and adolescence*, Springer, v. 49, n. 3, p. 618–630, 2020.

WEINSTEIN, R. S.; MADISON, S. M.; KUKLINSKI, M. R. Raising expectations in schooling: Obstacles and opportunities for change. *American Educational Research Journal*, Sage Publications, v. 32, n. 1, p. 121–159, 1995.

WELLINGS, K. et al. Sexual behaviour in context: a global perspective. *The Lancet*, Elsevier, v. 368, n. 9548, p. 1706–1728, 2006.

WILSON, K.; WOLFE, B.; HAVEMAN, R. The role of expectations in adolescent schooling choices: Do youths respond to economic incentives? *Economic Inquiry*, Wiley Online Library, v. 43, n. 3, p. 467–492, 2005.

WONG, J. S.; SCHONLAU, M. Does bully victimization predict future delinquency? a propensity score matching approach. *Criminal Justice and Behavior*, Sage Publications Sage CA: Los Angeles, CA, v. 40, n. 11, p. 1184–1208, 2013.

YOON, D. Peer-relationship patterns and their association with types of child abuse and adolescent risk behaviors among youth at-risk of maltreatment. *Journal of adolescence*, Elsevier, v. 80, p. 125–135, 2020.

ZAPPE, J. G.; ALVES, C. F.; AGLIO, D. D. D. Comportamentos de risco na adolescência: revisão sistemática de estudos empíricos. *Psicologia em Revista*, Pontifícia Universidade Católica de Minas Gerais, v. 24, n. 1, p. 79–100, 2018.

ZHANG, Y. Importance of home environment for children's schooling: From the teacher's perspective. In: *The impact and transformation of education policy in China*. [S.l.]: Emerald Group Publishing Limited, 2011.

APÊNDICE A – 1

TABLE 49 – Socioeconomic sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Missing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
Gender	102072	0					
Girls			52782	51.7	49784	52.4	0.7
Boys			49290	48.3	45263	47.6	-0.7
Age	102072	0					
11			28	0.0	14	0.0	0.0
12			510	0.5	479	0.5	0.0
13			16722	16.4	15756	16.6	0.2
14			51611	50.6	48370	50.9	0.3
15			20864	20.4	19250	20.3	-0.1
16			7873	7.7	7226	7.6	-0.1
17			3040	3.0	2726	2.9	-0.1
18			852	0.8	744	0.8	0.0
19			572	0.6	482	0.5	-0.1
City	102072	0					
Capital			51192	50.2	47849	50.3	0.1
Interior			50880	49.8	47198	49.7	-0.1
Region	102072	0					
North			23937	23.5	22258	23.4	-0.1
Northeast			36334	35.6	33876	35.6	0.0
Southeast			17772	17.4	16669	17.5	0.1
South			9850	9.7	9102	9.6	-0.1

Table 49 continued from previous page

Variable	Missing Data		Non-Missing Data				Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Midwest			14179	13.9	13142	13.8	-0.1
School type	102072	0					
Private			20918	20.5	19884	20.9	0.4
Public			81154	79.5	75163	79.1	-0.4
Color	101964	108					
White			33775	33.1	31576	33.2	0.1
Black			12849	12.6	11808	12.4	-0.2
Yellow			4580	4.5	4287	4.5	0.0
Brown-Skin			46935	46.0	43833	46.1	0.1
Indigenous			3825	3.8	3543	3.7	-0.1
Cellphone	101990	82					
Yes			88978	87.2	83169	87.5	0.3
No			13012	12.8	11878	12.5	-0.3
Internet	101967	105					
Yes			78395	76.9	73522	77.4	0.5
No			23572	23.1	21525	22.6	-0.5
Car	101955	117					
Yes			54845	53.8	51330	54.0	0.2
No			47110	46.2	43717	46.0	-0.2
Housekeeper	101976	96					
Yes			12096	11.9	11234	11.8	-0.1
No			89880	88.1	83813	88.2	0.1
Lives with mother	102001	71					

Table 49 continued from previous page

Variable	Missing Data		Non-Missing Data			Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Yes			90458	88.7	84309	88.7	0.0
No			11543	11.3	10738	11.3	0.0
Lives with father	101941	131					
Yes			63600	62.4	59274	62.4	0.0
No			38341	37.6	35773	37.6	0.0
People at home	102012	60					
1			162	0.2	130	0.1	-0.1
2			5898	5.8	5488	5.8	0.0
3			19555	19.2	18312	19.3	0.1
4			33174	32.5	31020	32.6	0.1
5			21925	21.5	20419	21.5	0.0
6			10452	10.2	9689	10.2	0.0
7			5029	4.9	4641	4.9	0.0
8 or more			5817	5.7	5348	5.6	-0.1

TABLE 50 – Relational/behavioral sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
At least one of the parents smoke	101780	292					
Yes			24362	23.9	22621	23.8	-0.1
No			77418	76.1	72426	76.2	0.1
Missed class without parental permission	101740	332					
Yes			21536	21.2	19937	21	-0.2
No			80204	78.8	75110	79	0.2
Body Perception	100995	1077					
Very Thin			5369	5.3	4946	5.2	-0.1
Thin			20896	20.7	19656	20.7	0.0
Normal			56611	56.1	53304	56.1	0.0
Fat			15919	15.8	15099	15.9	0.1
Very Fat			2200	2.2	2042	2.1	-0.1
Feels Lonely	101668	404					
Yes			17060	16.8	16126	17	0.2
No			84608	83.2	78921	83	-0.2
HPV Vaccination	101717	355					
Yes			37667	37	35868	37.7	0.7
No			64050	63	59179	62.3	-0.7
Parents Understand their Problems	101497	575					
Yes			43567	42.9	40933	43.1	0.2
No			57930	57.1	54114	56.9	-0.2

Table 50 continued from previous page

Variable	Missing Data		Non-Missing Data				Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Well-Treated by the Colleagues	101553	519					
Yes			63680	62.7	60284	63.4	0.7
No			37873	37.3	34763	36.6	-0.7
Practiced Vitimization	101597	475					
Yes			19002	18.7	17590	18.5	-0.2
No			82595	81.3	77457	81.5	0.2
Suffered Victimization	101505	567					
Yes			22612	22.3	20936	22	-0.3
No			78893	77.7	74111	78	0.3
Suffered Bullying	101677	395					
Yes			48117	47.3	45024	47.4	0.1
No			53560	52.7	50023	52.6	-0.1
Number of Close Friends	101638	434					
None			4179	4.1	3790	4	-0.1
1			6467	6.4	5996	6.3	-0.1
2			12972	12.8	12139	12.8	0.0
3 or more			78020	76.8	73122	76.9	0.1
Friends who drink alcohol	101824	248					
None or don’t know			29422	28.9	27255	28.7	-0.2
Few			28229	27.7	26302	27.7	0.0
Some			24067	23.6	22578	23.8	0.2
Most of			17290	17	16331	17.2	0.2
All			2816	2.8	2581	2.7	-0.1

Table 50 continued from previous page

Variable	Missing Data		Non-Missing Data				Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Friends who use Drugs	101791	281					
None or don't know			65998	64.8	61571	64.8	0.0
Few			18893	18.6	17721	18.6	0.0
Some			11642	11.4	10911	11.5	0.1
Most of			4517	4.4	4200	4.4	0.0
All			741	0.7	644	0.7	0.0
Victim of sexual violence	101300	772					
Yes			4124	4.1	3710	3.9	-0.2
No			97176	95.9	91337	96.1	0.2

TABLE 51 – Alcohol sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
Have you ever had alcohol	98717	3355					
Yes			51918	52.6	49823	52.4	-0.2
No			46799	47.4	45224	47.6	0.2
Age started drinking alcohol	101599	473					
Didn't drink			47505	46.8	45224	47.6	0.8
7 to 8			3050	3.0	2713	2.9	-0.1
9 to 10			4268	4.2	3885	4.1	-0.1
11 to 12			12159	12.0	11298	11.9	-0.1
13 to 14			27781	27.3	25763	27.1	-0.2
15 to 16			6164	6.1	5595	5.9	-0.2
17 to 18			672	0.7	569	0.6	-0.1
Frequency of alcohol use in the recent 30 days	101961	111					
Never took			47008	46.1	45224	47.6	1.5
none			32356	31.7	28619	30.1	-1.6
1 or 2 days			13773	13.5	12936	13.6	0.1
3 to 5 days			3905	3.8	3710	3.9	0.1
6 to 9 days			2176	2.1	2046	2.2	0.1
10 to 19 days			1487	1.5	1388	1.5	0.0
20 or more			1256	1.2	1124	1.2	0.0
Intensity of alcohol use in the recent 30 days	101918	154					
Never took			47008	46.1	45224	47.6	1.5

Table 51 continued from previous page

Variable	Missing Data		Non-Missing Data				Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Didn't take			29973	29.4	26357	27.7	-1.7
1 cup or less			12761	12.5	11882	12.5	0.0
2 to 3 cups			5834	5.7	5548	5.8	0.1
4 or more cups			6342	6.2	6036	6.4	0.2
Really been drunk in your life	101971	101					
Never took			47008	46.1	45224	47.6	1.5
None			34345	33.7	30510	32.1	-1.6
1 to 2 times			13593	13.3	12757	13.4	0.1
3 to 5 times			3633	3.6	3447	3.6	0.0
6 to 9 times			1152	1.1	1044	1.1	0.0
10 or more times			2240	2.2	2065	2.2	0.0
How purchased alcohol in the recent 30 days	101856	216					
Never took			47008	46.2	45224	47.6	1.4
I didn't take			25874	25.4	22508	23.7	-1.7
Shop or tavern			4131	4.1	3831	4.0	-0.1
Hawker			520	0.5	474	0.5	0.0
Someone bought for me			968	1.0	892	0.9	-0.1
Friends			5456	5.4	5173	5.4	0.0
Caught hiding			1221	1.2	1151	1.2	0.0
Older person gave it to me			2729	2.7	2560	2.7	0.0
Party			12363	12.1	11744	12.4	0.3
Achieved otherwise			1586	1.6	1490	1.6	0.0
Problems due to alcohol consumption	101950	122					

Table 51 continued from previous page

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Never took			47008	46.1	45224	47.6	1.5
None			47857	46.9	43472	45.7	-1.2
1 or 2 times in life			4430	4.3	4019	4.2	-0.1
3 to 5 times in life			1211	1.2	1085	1.1	-0.1
6 to 9 times in life			435	0.4	366	0.4	0.0
10 or more times in life			1009	1.0	881	0.9	-0.1

TABLE 52 – Cigarette sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
Smoked Cigarette Once	101881	191					
Yes			18723	18.4	17350	18.3	-0.1
No			83158	81.6	77697	81.7	0.1
Age started to smoke	101891	181					
Never Smoked			83349	81.8	77697	81.7	-0.1
7 to 8			1576	1.5	1427	1.5	0.0
9 to 10			1763	1.7	1629	1.7	0.0
11 to 12			4206	4.1	3969	4.2	0.1
13 to 14			8663	8.5	8164	8.6	0.1
15 to 16			2136	2.1	1988	2.1	0.0
17 to 17			198	0.2	173	0.2	0.0
Frequency use of cigarette in the recent 30 days	102013	59					
Never Smoked			83349	81.7	77697	81.7	0.0
None			13247	13.0	12345	13.0	0.0
1 or 2			2884	2.8	2686	2.8	0.0
3 to 5			858	0.8	799	0.8	0.0
6 to 9			465	0.5	431	0.5	0.0
10 to 19			374	0.4	357	0.4	0.0
20 or more			836	0.8	732	0.8	0.0
How did you acquired the cigarette in the recent 30 days	101968	104					
Never Smoked			83349	81.7	77697	81.7	0.0

Table 52 continued from previous page

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Did not Smoke			11075	10.9	10306	10.8	-0.1
Store or Bar			1848	1.8	1717	1.8	0.0
Hawker			428	0.4	392	0.4	0.0
Someone bought it for me			826	0.8	766	0.8	0.0
Asked someone			1488	1.5	1407	1.5	0.0
Got it hidden from someone			1198	1.2	1121	1.2	0.0
Got it from na older person			720	0.7	679	0.7	0.0
Got it in another way			1036	1.0	962	1.0	0.0
Somene refused to sell cigarettes	101963	109					
Never Smoked			83349	81.7	77697	81.7	0.0
Did not try to buy			13744	13.5	12844	13.5	0.0
Yes. they refused to sell to me			1637	1.6	1493	1.6	0.0
No. I bought it			3233	3.2	3013	3.2	0.0

TABLE 53 – Drug sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
Ever used drugs	101800	272					
Yes			8656	8.5	7882	8.3	-0.2
No			93144	91.5	87165	91.7	0.2
Age started to use drugs	101850	222					
Never Used			93416	91.7	87165	91.7	0.0
7 to 8			247	0.2	198	0.2	0.0
9 to 10			230	0.2	208	0.2	0.0
11 to 12			1095	1.1	1027	1.1	0.0
13 to 14			4982	4.9	4688	4.9	0.0
15 to 16			1736	1.7	1632	1.7	0.0
17 to 18			144	0.1	129	0.1	0.0
Frequency use of drugs in the recent 30 days	102030	42					
Never Used			93416	91.6	87165	91.7	0.1
None			4699	4.6	4231	4.5	-0.1
1 or 2			1992	2.0	1874	2.0	0.0
3 to 5			729	0.7	690	0.7	0.0
6 to 9			365	0.4	341	0.4	0.0
10 or more			829	0.8	746	0.8	0.0
Frequency use of marijuana in the recent 30 days	102032	40					
Never Used			93416	91.6	87165	91.7	0.1
None			4754	4.7	4286	4.5	-0.2

Table 53 continued from previous page

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
1 or 2			2080	2.0	1955	2.1	0.1
3 to 9			811	0.8	761	0.8	0.0
10 or more			971	1.0	880	0.9	-0.1
Frequency use of Crack in the recent 30 days	102006	66					
Never Used			93416	91.6	87165	91.7	0.1
None			8070	7.9	7461	7.8	-0.1
1 or 2			212	0.2	173	0.2	0.0
3 to 9			130	0.1	102	0.1	0.0
10 or more			178	0.2	146	0.2	0.0

TABLE 54 – Sexual risk sample data - PeNSE

Variable	Missing Data				Non-Missing Data		Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
Total	102072				95047	100	
Ever had sexual intercourse	101566	506					
Yes			28577	28.1	26301	27.7	-0.4
No			72989	71.9	68746	72.3	0.4
Age that had the first sexual relation	101841	231					
Never had it			73495	72.2	68746	72.3	0.1
9 to 10			2973	2.9	2665	2.8	-0.1
11 to 12			4436	4.4	4115	4.3	-0.1
13 to 14			15460	15.2	14446	15.2	0.0
15 to 16			5031	4.9	4677	4.9	0.0
17 to 18			446	0.4	398	0.4	0.0
Number of sexual partners	101896	176					
Never had it			73495	72.1	68746	72.3	0.2
1			10218	10.0	9493	10.0	0.0
2			5483	5.4	5134	5.4	0.0
3			3681	3.6	3413	3.6	0.0
4			2145	2.1	1983	2.1	0.0
5			1506	1.5	1387	1.5	0.0
6 or more			5368	5.3	4891	5.1	-0.2
Used preservative in the first sexual relation	101961	111					
Never had it			73495	72.1	68746	72.3	0.2
Yes			17465	17.1	16116	17.0	-0.1

Table 54 continued from previous page

Variable	Missing Data		Non-Missing Data				Difference
	N	Misssing	Frequency	% (A)	Frequency	% (B)	(B) - (A)
No			11001	10.8	10185	10.7	-0.1
Used preservative in the last sexual relation	101899	173					
Never had it			73495	72.1	68746	72.3	0.2
Yes			18960	18.6	17557	18.5	-0.1
No			8446	8.3	7889	8.3	0.0
Do not recall			998	1.0	855	0.9	-0.1

TABLE 55 – F and Qui-square test for individual significance - Table 13

Variable	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Gender	1	18.93	<0.01	18.93	<0.01
Capital	1	46.10	<0.01	46.10	<0.01
Region	4	43.70	<0.01	174.81	<0.01
Public	1	25.45	<0.01	25.45	<0.01
Color	4	2.88	0.021	11.52	0.021
Cellphone	1	218.29	<0.01	218.29	<0.01
Internet	1	69.97	<0.01	69.97	<0.01
Car	1	5.04	0.025	5.04	0.025
Housekeeper	1	7.56	<0.01	7.56	<0.01
Lives with father	1	0.57	0.450	0.57	0.450
Lives with mother	1	139.65	<0.01	139.65	<0.01
People at home	1	1226.52	<0.01	1226.52	<0.01

TABLE 56 – F and Qui-square test for individual significance - Table 14

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Parents smoke	1	88.18	<0.01	88.18	<0.01
Miss classes	1	9.71	<0.01	9.71	<0.01
Body perception	1	37.86	<0.01	151.42	<0.01
Feel alone	1	39.99	<0.01	39.99	<0.01
HPV	4	6.65	<0.01	6.65	<0.01
Comprehensive parents	1	284.07	<0.01	284.07	<0.01
Colleagues treat well	1	22.99	<0.01	22.99	<0.01
Suffered victimization	1	2.13	0.145	2.13	0.145
Practiced Victimization	1	77.87	<0.01	77.87	<0.01
Bullying	1	0.54	0.462	0.54	0.461
Close friends	1	6.41	0.011	6.41	0.011
Friends drink alcohol	1	6.46	0.011	6.46	0.011
Friends use drugs	1	1.09	0.297	1.09	0.297
Sexual violence	1	537.98	<0.01	537.98	<0.01

TABLE 57 – F and Qui-square test for individual significance - Table 15

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Already smoked cigarette	1	7.82	<0.01	7.82	<0.01
Already drank alcohol	1	0.05	0.823	0.05	0.823
Have used drugs	1	48.23	<0.01	48.23	<0.01
Have had sex	1	2835.88	<0.01	2835.88	<0.01

TABLE 58 – F and Qui-square test for individual significance - Table 16

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Cigarette use 30 days	6	5.75	<0.01	34.49	<0.01
Alcohol use 30 days	6	4.87	<0.01	29.22	<0.01
Drug use 30 days	5	2.40	0.04	12.00	0.035
Sex Partners	6	495.86	<0.01	2975.17	<0.01

TABLE 59 – F and Qui-square test for individual significance - Table 17

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Gender	1	15.41	<0.01	15.41	<0.01
Capital	1	0.25	0.616	0.25	0.616
Region	4	4.39	<0.01	17.58	<0.01
Public	1	18.74	<0.01	18.74	<0.01
Color	4	2.64	0.032	10.58	0.032
Cellphone	1	7.64	<0.01	7.64	<0.01
Internet	1	39.97	<0.01	39.97	<0.01
Car	1	2.17	0.141	2.17	0.140
Housekeeper	1	2.90	0.089	2.90	0.089
Lives with father	1	1.54	0.215	1.54	0.215
Lives with mother	1	0.18	0.668	0.18	0.668
People at House	1	18.49	<0.01	18.49	<0.01

TABLE 60 – F and Qui-square test for individual significance - Table 18

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Parents smoke	1	2.62	0.106	2.62	0.106
Miss classes	1	3.34	0.068	3.34	0.068
Body perception	4	0.66	0.622	2.63	0.622
Feel alone	1	0.62	0.430	0.62	0.430
HPV	1	9.54	<001	9.54	<001
Comprehensive parents	1	27.75	<001	27.75	<001
Colleagues treat well	1	0.70	0.403	0.70	0.403
Suffered victimization	1	51.78	<001	51.78	<001
Practiced Victimization	1	1.45	0.229	1.45	0.229
Bullying	1	0.35	0.552	0.35	0.552
Close friends	1	19.68	<001	19.68	<001
Friends drink alcohol	1	0.08	0.777	0.08	0.777
Friends use drugs	1	25.83	<001	25.83	<001
Sexual violence	1	0.01	0.934	0.01	0.934

TABLE 61 – F and Qui-square test for individual significance - Table 19

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Already smoked cigarette	1	8.01	<001	8.01	<001
Already drank alcohol	1	1.15	0.283	1.15	0.283
Have used drugs	1	1.43	0.232	1.43	0.232
Have had sex	1	6.13	0.013	6.13	0.013

TABLE 62 – F and Qui-square test for individual significance - Table 20

Variables	Df	F	Pr(>F)	Chisq	Pr(>Chisq)
Cigarette use 30 days	6	2.78	0.011	16.66	0.011
Alcohol use 30 days	6	1.89	0.079	11.35	0.078
Drug use 30 days	5	1.00	0.415	5.01	0.414
Sex Partners	6	2.53	0.019	15.18	0.019

APÊNDICE B – 2

TABLE 63 – Fit indices for different exploratory factor analysis estimator - No rotation applied

	minres	gls	wls	ml
TLI	0.89	0.89	0.90	0.90
RMSR	0.03	0.03	0.03	0.03
RMSEA	0.13	0.13	0.13	0.12
Chi-square	69330.77	66277.67	63297.19	60340.82
BIC	68745.62	65692.51	62712.03	59755.67

TABLE 64 – MLM estimation with no rotation

	ML1	ML2	ML3
agealc	0.354		0.635
alc_freq30	0.552		0.677
alc_int30	0.538		0.641
drunk	0.588		0.579
probeb	0.445		0.586
cigar30	0.631		
agedrug	0.816		
drug30	0.895	-0.331	
marijuana30	0.919	-0.339	
crack30	0.827		
agesex	0.515	0.69	
preserv	0.559	0.748	
preserv1	0.574	0.759	
sexpartners	0.548	0.551	
SS loadings	5.865	2.474	2.05
Proportion Var	0.419	0.177	0.146
Cumulative Var	0.419	0.596	0.742

TABLE 65 – MLM estimation with promax rotation

	ML1	ML3	ML2
agealc		0.837	
alc_freq30		0.89	
alc_int30		0.846	
drunk		0.776	
probeb		0.773	
cigar30	0.487		
agedrug	0.849		
drug30	1.004		
marijuana30	1.032		
crack30	0.879		
agesex			0.876
preserv			0.975
preserv1			0.991
sexpartners			0.729
SS loadings	3.852	3.467	3.241
Proportion Var	0.275	0.248	0.232
Cumulative Var	0.275	0.523	0.754

TABLE 66 – MLM estimation with varimax rotation

	ML1	ML3	ML2
agealc		0.755	
alc_freq30		0.842	
alc_int30		0.804	
drunk		0.765	
probeb		0.724	
cigar30	0.526	0.323	
agedrug	0.812		
drug30	0.936		
marijuana30	0.962		
crack30	0.835		
agesex			0.835
preserv			0.92
preserv1			0.936
sexpartners			0.719
SS loadings	3.759	3.418	3.211
Proportion Var	0.269	0.244	0.229
Cumulative Var	0.269	0.513	0.742

TABLE 67 – MLM estimation with oblimin rotation

	ML1	ML3	ML2
agealc		0.821	
alc_freq30		0.884	
alc_int30		0.84	
drunk		0.775	
probeb		0.765	
cigar30	0.473		
agedrug	0.821		
drug30	0.969		
marijuana30	0.996		
crack30	0.849		
agesex			0.863
preserv			0.96
preserv1			0.976
sexpartners			0.721
SS loadings	3.599	3.409	3.15
Proportion Var	0.257	0.243	0.225
Cumulative Var	0.257	0.501	0.726

TABLE 68 – Hierarchical - Schmid Leiman Factor loadings greater than 0.2

	g	F1*	F2*	F3*	h2	u2	p2
agealc	0.56		0.56		0.64	0.36	0.49
alc_freq30	0.67		0.56		0.77	0.23	0.59
alc_int30	0.64		0.52		0.68	0.32	0.6
drunk	0.68		0.52		0.75	0.25	0.63
probeb	0.58		0.52		0.61	0.39	0.56
cigar30	0.48	0.39			0.41	0.59	0.57
agedrug	0.52	0.71			0.77	0.23	0.35
drug30	0.52	0.78			0.87	0.13	0.31
marijuana30	0.54	0.81			0.94	0.06	0.31
crack30	0.52	0.74			0.81	0.19	0.33
agesex	0.51			0.71	0.76	0.24	0.34
preserv	0.53			0.78	0.89	0.11	0.32
preserv1	0.55			0.80	0.93	0.07	0.32
sexpartners	0.51			0.59	0.61	0.39	0.42
	g	F1*	F2*	F3*			
Eigenvalues Hierarchical	4.4	2.5	1.4	2.1			

Explained Common Variance of the general factor = 0.42

TABLE 69 – Fit Indices - Hierarchical

RMSR	0.03
RMSEA	0.131
Chi-square	69330.77
BIC	68745.62

TABLE 70 – Fit indices SEM models - Test for Svydesign

	Sample	Weights Applied	diff
Robust CFI	0.925	0.924	-0.001
Robust TLI	0.907	0.906	-0.001
GFI	0.88	0.958	0.078
RMSR	0.063	0.058	-0.005
RMSEA	0.11	0.11	0
Chi-square	79793502	80679172	885670
BIC	2218660142	2249626682	30966540
AIC	2218336289	2249164035	30827746

TABLE 71 – Correlation variable matrix

	agesex	preserv	preserv1	sexpartn	agedrug	mariju30	crack30	drug30	cigar30	agealc	alc_freq30	alc_int30	drunk	probeb	schooexp
agesex	1.000														
preserv	0.823	1.000													
preserv1	0.839	0.913	1.000												
sexpartn	0.673	0.733	0.746	1.000											
agedrug	0.241	0.263	0.267	0.215	1.000										
mariju30	0.278	0.303	0.309	0.248	0.833	1.000									
crack30	0.246	0.267	0.272	0.219	0.736	0.849	1.000								
drug30	0.271	0.295	0.301	0.241	0.812	0.937	0.828	1.000							
cigar30	0.237	0.258	0.263	0.211	0.498	0.574	0.507	0.560	1.000						
agealc	0.281	0.305	0.311	0.250	0.281	0.324	0.286	0.316	0.357	1.000					
alc_freq30	0.341	0.371	0.378	0.303	0.341	0.393	0.347	0.383	0.433	0.655	1.000				
alc_int30	0.331	0.360	0.367	0.295	0.331	0.382	0.337	0.372	0.421	0.636	0.772	1.000			
drunk	0.321	0.349	0.356	0.285	0.321	0.370	0.327	0.361	0.408	0.616	0.748	0.727	1.000		
probeb	0.284	0.309	0.315	0.253	0.284	0.328	0.289	0.319	0.361	0.546	0.662	0.644	0.624	1.000	
schooexp	-0.083	-0.090	-0.091	-0.073	-0.025	-0.029	-0.025	-0.028	-0.022	-0.022	-0.027	-0.026	-0.026	-0.023	1.000

TABLE 72 – Latent Variables for estimates multi-group by sex

		Men				Women			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
rsex =									
agersx		1.000				1.000			
preserv	(.p2.)	0.512	0.001	434.850	0.000	0.512	0.001	434.850	0.000
presrv1	(.p3.)	0.519	0.001	451.826	0.000	0.519	0.001	451.826	0.000
sexpartn	(.p4.)	0.906	0.003	278.921	0.000	0.906	0.003	278.921	0.000
drug =									
agedrug		1.000				1.000			
marij30	(.p6.)	0.576	0.001	418.375	0.000	0.576	0.001	418.375	0.000
crack30	(.p7.)	0.299	0.001	331.455	0.000	0.299	0.001	331.455	0.000
drug30	(.p8.)	0.636	0.002	397.625	0.000	0.636	0.002	397.625	0.000
cigar30	(.p9.)	0.530	0.003	185.470	0.000	0.530	0.003	185.470	0.000
alc =									
agealc		1.000				1.000			
alc_f30	(.11.)	0.769	0.003	256.244	0.000	0.769	0.003	256.244	0.000
alc_in30	(.12.)	1.121	0.005	245.948	0.000	1.121	0.005	245.948	0.000
drunk	(.13.)	0.634	0.003	239.647	0.000	0.634	0.003	239.647	0.000
probeb	(.14.)	0.408	0.002	211.478	0.000	0.408	0.002	211.478	0.000
schooexp=									
maxscho		1.000				1.000			

TABLE 73 – Variances for estimates multi-group by sex

		Men				Women			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
agesex		0.701	0.006	122.211	0.000	0.275	0.002	121.543	0.000
.preserv		0.073	0.001	100.133	0.000	0.026	0.000	80.946	0.000
.presrv1	(.21.)	0.027	0.000	89.446	0.000	0.027	0.000	89.446	0.000
sexpartn		1.700	0.013	130.733	0.000	0.489	0.004	133.144	0.000
agedrug	(.23.)	0.351	0.002	183.422	0.000	0.351	0.002	183.422	0.000
mariju30	(.24.)	0.012	0.000	71.483	0.000	0.012	0.000	71.483	0.000
.crack30		0.038	0.000	127.633	0.000	0.017	0.000	125.990	0.000
drug30	(.26.)	0.035	0.000	129.254	0.000	0.035	0.000	129.254	0.000
.cigar30	(.27.)	0.468	0.002	193.770	0.000	0.468	0.002	193.770	0.000
agealc	(.28.)	1.664	0.009	176.131	0.000	1.664	0.009	176.131	0.000
.alc_f30	(.29.)	0.302	0.002	129.017	0.000	0.302	0.002	129.017	0.000
.alc_in30	(.30.)	0.906	0.006	148.471	0.000	0.906	0.006	148.471	0.000
.drunk	(.31.)	0.344	0.002	156.322	0.000	0.344	0.002	156.322	0.000
.probeb	(.32.)	0.266	0.002	175.333	0.000	0.266	0.002	175.333	0.000
schooexp		0.000				0.000			
Sexual		1.897	0.016	116.840	0.000	1.136	0.009	123.690	0.000
Drugs		1.107	0.010	114.562	0.000	0.757	0.006	118.652	0.000
Alcohol		2.159	0.022	97.571	0.000	2.020	0.020	100.596	0.000
Expectation		1.937	0.014	134.864	0.000	1.741	0.012	142.652	0.000

TABLE 74 – Latent Variables for estimates multi-group by area

	Capital				Interior			
	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
rsex =								
agersx	1.000				1.000			
preserv	0.550	0.002	293.639	0.000	0.524	0.002	277.560	0.000
presrv1	0.555	0.002	304.523	0.000	0.538	0.002	288.056	0.000
sexpartn	1.047	0.005	194.801	0.000	1.055	0.006	189.140	0.000
drug =								
agedrug	1.000				1.000			
marij30	0.591	0.002	296.568	0.000	0.559	0.002	290.906	0.000
crack30	0.296	0.001	230.976	0.000	0.303	0.001	227.929	0.000
drug30	0.650	0.002	282.113	0.000	0.621	0.002	276.637	0.000
cigar30	0.533	0.004	138.992	0.000	0.530	0.004	121.679	0.000
alc =								
agealc	1.000				1.000			
alc_f30	0.773	0.004	184.724	0.000	0.767	0.004	177.518	0.000
alc_in30	1.148	0.006	176.707	0.000	1.095	0.006	170.862	0.000
drunk	0.644	0.004	170.935	0.000	0.625	0.004	167.820	0.000
probeb	0.408	0.003	149.411	0.000	0.408	0.003	149.504	0.000
schooexp=								
maxscho	1.000				1.000			

TABLE 75 – Latent Variables for estimates multi-group by school

		Public				Private			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
rsex =									
agersx		1.000				1.000			
preserv	(.p2.)	0.539	0.001	403.229	0.000	0.539	0.001	403.229	0.000
presrv1	(.p3.)	0.549	0.001	422.025	0.000	0.549	0.001	422.025	0.000
sexpartn	(.p4.)	1042	0.004	271.797	0.000	1042	0.004	271.797	0.000
drug =									
agedrug		1.000				1.000			
marij30	(.p6.)	0.577	0.001	415.622	0.000	0.577	0.001	415.622	0.000
crack30	(.p7.)	0.299	0.001	324.229	0.000	0.299	0.001	324.229	0.000
drug30	(.p8.)	0.638	0.002	395.374	0.000	0.638	0.002	395.374	0.000
cigar30	(.p9.)	0.531	0.003	185.228	0.000	0.531	0.003	185.228	0.000
alc =									
agealc		1.000				1.000			
alc_f30	(.11.)	0.770	0.003	256.071	0.000	0.770	0.003	256.071	0.000
alc_in30	(.12.)	1.122	0.005	245.718	0.000	1.122	0.005	245.718	0.000
drunk	(.13.)	0.635	0.003	239.442	0.000	0.635	0.003	239.442	0.000
probeb	(.14.)	0.408	0.002	211.222	0.000	0.408	0.002	211.222	0.000
schooexp=									
maxscho		1.000				1.000			

TABLE 76 – Variances for estimates multi-group by school

		Public				Private			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
agesex		0.462	0.003	174.635	0.000	0.462	0.003	174.635	0.000
.preserv		0.057	0.000	116.686	0.000	0.017	0.000	41.154	0.000
.presrv1	(.21.)	0.029	0.000	84.775	0.000	0.029	0.000	84.775	0.000
sexpartn		1.175	0.007	164.866	0.000	0.555	0.006	86.229	0.000
agedrug	(.23.)	0.355	0.002	183.575	0.000	0.355	0.002	183.575	0.000
mariju30	(.24.)	0.012	0.000	68.016	0.000	0.012	0.000	68.016	0.000
.crack30		0.027	0.000	181.102	0.000	0.027	0.000	181.102	0.000
drug30	(.26.)	0.034	0.000	126.835	0.000	0.034	0.000	126.835	0.000
.cigar30	(.27.)	0.468	0.002	193.776	0.000	0.468	0.002	193.776	0.000
agealc	(.28.)	1.666	0.009	176.183	0.000	1.666	0.009	176.183	0.000
.alc_f30	(.29.)	0.301	0.002	128.724	0.000	0.301	0.002	128.724	0.000
.alc_in30	(.30.)	0.905	0.006	148.431	0.000	0.905	0.006	148.431	0.000
.drunk	(.31.)	0.344	0.002	156.286	0.000	0.344	0.002	156.286	0.000
.probeb	(.32.)	0.267	0.002	175.370	0.000	0.267	0.002	175.370	0.000
schooexp		0.000				0.000			
Sexual		1.494	0.011	139.290	0.000	0.881	0.011	83.014	0.000
Drugs		0.971	0.007	135.979	0.000	0.734	0.009	83.121	0.000
Alcohol		2.118	0.019	111.879	0.000	1.946	0.026	74.250	0.000
Expectation		1.976	0.011	173.468	0.000	1.105	0.012	91.915	0.000

TABLE 77 – Latent Variables for estimates multi-group by region

		Midwest				South				Southeast				Notheast				North			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
rsex =																					
agersx		1.000				1.000				1.000				1.000				1.000			
preserv	(.p2.)	0.537	0.001	402.933	0.000	0.537	0.001	402.933	0.000	0.537	0.001	402.933	0.000	0.537	0.001	402.933	0.000	0.537	0.001	402.933	0.000
presrv1	(.p3.)	0.547	0.001	418.606	0.000	0.547	0.001	418.606	0.000	0.547	0.001	418.606	0.000	0.547	0.001	418.606	0.000	0.547	0.001	418.606	0.000
sexpartn	(.p4.)	1052	0.004	271.637	0.000	1052	0.004	271.637	0.000	1052	0.004	271.637	0.000	1052	0.004	271.637	0.000	1052	0.004	271.637	0.000
drug =																					
agedrug		1.000				1.000				1.000				1.000				1.000			
marij30	(.p6.)	0.566	0.001	421.167	0.000	0.566	0.001	421.167	0.000	0.566	0.001	421.167	0.000	0.566	0.001	421.167	0.000	0.566	0.001	421.167	0.000
crack30	(.p7.)	0.297	0.001	330.768	0.000	0.297	0.001	330.768	0.000	0.297	0.001	330.768	0.000	0.297	0.001	330.768	0.000	0.297	0.001	330.768	0.000
drug30	(.p8.)	0.626	0.002	400.750	0.000	0.626	0.002	400.750	0.000	0.626	0.002	400.750	0.000	0.626	0.002	400.750	0.000	0.626	0.002	400.750	0.000
cigar30	(.p9.)	0.520	0.003	185.923	0.000	0.520	0.003	185.923	0.000	0.520	0.003	185.923	0.000	0.520	0.003	185.923	0.000	0.520	0.003	185.923	0.000
alc =																					
agealc		1.000				1.000				1.000				1.000				1.000			
alc_f30	(.11.)	0.771	0.003	255.985	0.000	0.771	0.003	255.985	0.000	0.771	0.003	255.985	0.000	0.771	0.003	255.985	0.000	0.771	0.003	255.985	0.000
alc_in30	(.12.)	1.123	0.005	245.650	0.000	1.123	0.005	245.650	0.000	1.123	0.005	245.650	0.000	1.123	0.005	245.650	0.000	1.123	0.005	245.650	0.000
drunk	(.13.)	0.635	0.003	239.278	0.000	0.635	0.003	239.278	0.000	0.635	0.003	239.278	0.000	0.635	0.003	239.278	0.000	0.635	0.003	239.278	0.000
probeb	(.14.)	0.408	0.002	211.022	0.000	0.408	0.002	211.022	0.000	0.408	0.002	211.022	0.000	0.408	0.002	211.022	0.000	0.408	0.002	211.022	0.000
schooexp=																					
maxscho		1.000				1.000				1.000				1.000				1.000			

TABLE 78 – Variances for multi-group estimates by region

		Midwest				South				Southeast				Notheast				North			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
agesex		0.459	0.003	172.667	0.000	0.459	0.003	172.667	0.000	0.459	0.003	172.667	0.000	0.459	0.003	172.667	0.000	0.459	0.003	172.667	0.000
.preserv		0.047	0.000	117.703	0.000	0.047	0.000	117.703	0.000	0.047	0.000	117.703	0.000	0.047	0.000	117.703	0.000	0.047	0.000	117.703	0.000
.presrv1	(.21.)	0.031	0.000	86.109	0.000	0.031	0.000	86.109	0.000	0.031	0.000	86.109	0.000	0.031	0.000	86.109	0.000	0.031	0.000	86.109	0.000
sexpartn		1.033	0.006	185.072	0.000	1.033	0.006	185.072	0.000	1.033	0.006	185.072	0.000	1.033	0.006	185.072	0.000	1.033	0.006	185.072	0.000
agedrug	(.23.)	0.497	0.007	69.783	0.000	0.633	0.011	57.973	0.000	0.384	0.005	77.384	0.000	0.225	0.002	106.492	0.000	0.332	0.004	88.466	0.000
mariju30	(.24.)	0.012	0.000	74.026	0.000	0.012	0.000	74.026	0.000	0.012	0.000	74.026	0.000	0.012	0.000	74.026	0.000	0.012	0.000	74.026	0.000
.crack30		0.040	0.001	69.497	0.000	0.051	0.001	57.751	0.000	0.027	0.000	76.194	0.000	0.017	0.000	104.351	0.000	0.024	0.000	87.167	0.000
drug30	(.26.)	0.034	0.000	130.416	0.000	0.034	0.000	130.416	0.000	0.034	0.000	130.416	0.000	0.034	0.000	130.416	0.000	0.034	0.000	130.416	0.000
.cigar30	(.27.)	0.657	0.009	72.560	0.000	0.727	0.012	59.731	0.000	0.414	0.005	81.024	0.000	0.302	0.003	115.296	0.000	0.551	0.006	93.833	0.000
agealc	(.28.)	1.669	0.009	176.288	0.000	1.669	0.009	176.288	0.000	1.669	0.009	176.288	0.000	1.669	0.009	176.288	0.000	1.669	0.009	176.288	0.000
.alc_f30	(.29.)	0.300	0.002	128.738	0.000	0.300	0.002	128.738	0.000	0.300	0.002	128.738	0.000	0.300	0.002	128.738	0.000	0.300	0.002	128.738	0.000
.alc_in30	(.30.)	0.904	0.006	148.462	0.000	0.904	0.006	148.462	0.000	0.904	0.006	148.462	0.000	0.904	0.006	148.462	0.000	0.904	0.006	148.462	0.000
.drunk	(.31.)	0.345	0.002	156.436	0.000	0.345	0.002	156.436	0.000	0.345	0.002	156.436	0.000	0.345	0.002	156.436	0.000	0.345	0.002	156.436	0.000
.probeb	(.32.)	0.267	0.002	175.476	0.000	0.267	0.002	175.476	0.000	0.267	0.002	175.476	0.000	0.267	0.002	175.476	0.000	0.267	0.002	175.476	0.000
schooexp		0.000				0.000				0.000				0.000				0.000			
Sexual		1.370	0.020	67.913	0.000	1.236	0.022	56.328	0.000	1.270	0.017	75.302	0.000	1.311	0.013	102.933	0.000	1.620	0.019	86.247	0.000
Drugs		1.273	0.019	68.568	0.000	1678	0.029	57.435	0.000	1076	0.014	76.007	0.000	0.612	0.006	102.570	0.000	0.870	0.010	85.915	0.000
Alcohol		2.310	0.037	62.450	0.000	2.356	0.045	52.821	0.000	2.199	0.032	68.480	0.000	1.905	0.022	88.601	0.000	1.902	0.025	75.937	0.000
Expectation		1.736	0.024	73.233	0.000	1.669	0.028	60.241	0.000	1.694	0.021	82.185	0.000	1.955	0.017	117.410	0.000	1.884	0.020	94.835	0.000

TABLE 79 – Latent Variables for estimates multi-group by region

		Native				Yellow				Black				Brown/Mixed				White			
		Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)	Estimate	Std.Err	z-value	P(>|z|)
rsex =																					
agersx		1.000				1.000				1.000				1.000				1.000			
preserv	(.p2.)	0.556	0.008	69.558	0.000	0.514	0.006	87.026	0.000	0.553	0.004	131.175	0.000	0.539	0.002	273.549	0.000	0.525	0.002	243.415	0.000
presrv1	(.p3.)	0.593	0.008	73.725	0.000	0.535	0.006	92.113	0.000	0.565	0.004	136.232	0.000	0.549	0.002	282.446	0.000	0.529	0.002	252.986	0.000
sexpartn	(.p4.)	1161	0.023	50.364	0.000	1008	0.017	58.305	0.000	1113	0.012	90.302	0.000	1049	0.006	183.506	0.000	1013	0.006	160.459	0.000
drug =																					
agedrug		1.000				1.000				1.000				1.000				1.000			
marij30	(.p6.)	0.643	0.009	73.653	0.000	0.634	0.008	78.596	0.000	0.594	0.004	135.098	0.000	0.547	0.002	297.317	0.000	0.593	0.002	241.853	0.000
crack30	(.p7.)	0.363	0.006	56.940	0.000	0.336	0.005	62.180	0.000	0.307	0.003	106.720	0.000	0.280	0.001	233.230	0.000	0.307	0.002	187.264	0.000
drug30	(.p8.)	0.711	0.010	70.477	0.000	0.727	0.010	75.783	0.000	0.670	0.005	129.942	0.000	0.600	0.002	281.966	0.000	0.650	0.003	228.838	0.000
cigar30	(.p9.)	0.662	0.018	36.837	0.000	0.606	0.014	42.039	0.000	0.564	0.009	64.708	0.000	0.506	0.004	125.632	0.000	0.520	0.005	105.570	0.000
alc =																					
agealc		1.000				1.000				1.000				1.000				1.000			
alc_f30	(.11.)	0.797	0.017	46.059	0.000	0.800	0.015	52.157	0.000	0.801	0.009	85.750	0.000	0.725	0.004	179.873	0.000	0.812	0.006	145.989	0.000
alc_in30	(.12.)	1.174	0.026	44.968	0.000	1.170	0.024	49.508	0.000	1.159	0.014	82.376	0.000	1.061	0.006	171.136	0.000	1.178	0.008	141.296	0.000
drunk	(.13.)	0.675	0.015	43.821	0.000	0.645	0.013	48.338	0.000	0.659	0.008	80.144	0.000	0.604	0.004	168.321	0.000	0.662	0.005	136.444	0.000
probeb	(.14.)	0.457	0.012	38.692	0.000	0.422	0.010	43.216	0.000	0.416	0.006	71.005	0.000	0.395	0.003	148.445	0.000	0.416	0.003	120.143	0.000
schooexp=																					
maxscho		1.000				1.000				1.000				1.000				1.000			

TABLE 80 – Latent Variables for estimates with bullying mediation

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
rsex =						
agersx	1.000				1180	0.867
preserv	0.537	0.002	234010	0.000	0.633	0.947
presrv1	0.547	0.002	235406	0.000	0.645	0.964
sexpartn	1051	0.006	165160	0.000	1240	0.773
drug =						
agedrug	1.000				0.959	0.849
marij30	0.578	0.006	100798	0.000	0.554	0.982
crack30	0.299	0.003	88680	0.000	0.287	0.867
drug30	0.638	0.007	90314	0.000	0.612	0.957
cigar30	0.417	0.009	45090	0.000	0.400	0.468
alc =						
agealc	1.000				1436	0.742
alc_f30	0.773	0.005	145388	0.000	1111	0.896
alc_in30	1128	0.007	158087	0.000	1620	0.863
drunk	0.639	0.004	148784	0.000	0.919	0.844
probeb	0.410	0.003	142034	0.000	0.589	0.752
cigar30	0.166	0.003	48461	0.000	0.239	0.280
schooexp =						
maxschoo	1.000				1363	1000
g =						
Alcohol	1.152	0.006	203128	0.000	0.802	0.802
Sexual	0.622	0.005	121693	0.000	0.528	0.528
Drugs	0.551	0.006	87173	0.000	0.575	0.575
Bullying =						
feelalone	1.000				0.109	0.290
treatwell	-1.648	0.040	-40881	0.000	-0.179	-0.375
victimized	2250	0.053	42152	0.000	0.245	0.592
bullying	1909	0.044	43639	0.000	0.208	0.416
peereffects =						
drugfriends	1.000				0.647	0.717
alcoholfriends	1.204	0.010	116917	0.000	0.779	0.683

TABLE 81 – Latent Variables for estimates with peer effect mediation

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
rsex =						
agersx	1.000				1179	0.867
preserv	0.537	0.002	234044	0.000	0.633	0.947
presrv1	0.547	0.002	235396	0.000	0.645	0.964
sexpartn	1051	0.006	165152	0.000	1240	0.773
drug =						
agedrug	1.000				0.959	0.849
marij30	0.578	0.006	100795	0.000	0.554	0.981
crack30	0.299	0.003	88677	0.000	0.287	0.867
drug30	0.638	0.007	90314	0.000	0.612	0.957
cigar30	0.417	0.009	45186	0.000	0.400	0.468
alc =						
agealc	1.000				1437	0.742
alc_f30	0.773	0.005	145413	0.000	1111	0.896
alc_in30	1127	0.007	158110	0.000	1620	0.862
drunk	0.639	0.004	148846	0.000	0.919	0.844
probeb	0.410	0.003	142051	0.000	0.589	0.752
cigar30	0.166	0.003	48665	0.000	0.239	0.280
schooexp =						
maxschoo	1.000				1363	1000
g =						
Alcohol	1.148	0.006	204562	0.000	0.799	0.799
Sexual	0.623	0.005	122254	0.000	0.528	0.528
Drugs	0.549	0.006	87201	0.000	0.573	0.573
Bullying =						
feelalone	1.000				0.123	0.327
treatwell	-1.382	0.032	-42729	0.000	-0.170	-0.356
victimized	1851	0.040	46391	0.000	0.228	0.550
bullying	1822	0.039	46693	0.000	0.224	0.448
peereffects =						
drugfriends	1.000				0.648	0.717
alcoholfriends	1.202	0.010	117669	0.000	0.778	0.683

TABLE 82 – Fit indices for estimates with cigar in both - Table 24

	Robust CFI	Robust TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.930	0.915	0.960	0.045	0.105	74363044	2243299301	2242845907
Model 2	0.930	0.915	0.960	0.046	0.105	74526860	2243463117	2243009722
Model 3	0.931	0.915	0.961	0.045	0.105	73940298	2242876555	2242423161
Model 4	0.930	0.915	0.960	0.047	0.105	74541772	2243478029	2243024635
Model 5	0.930	0.914	0.960	0.045	0.106	74171340	2243118850	2242656202
Model 6	0.931	0.914	0.961	0.045	0.106	73932560	2242880070	2242417422
Model 7	0.930	0.914	0.960	0.045	0.106	74346639	2243294149	2242831502
Model 8	0.931	0.914	0.961	0.045	0.106	73929527	2242877037	2242414390
Model 9	0.930	0.914	0.960	0.046	0.106	74506965	2243454475	2242991828
Model 10	0.931	0.914	0.961	0.045	0.106	73940064	2242887574	2242424926
Model 11	0.931	0.913	0.961	0.045	0.106	73928849	2242887612	2242415711
Model 12	0.931	0.913	0.961	0.045	0.106	73928849	2242887612	2242415711
Model 13	0.931	0.913	0.961	0.045	0.106	73928849	2242887612	2242415711
Model 14	0.931	0.913	0.961	0.045	0.106	73928849	2242887612	2242415711

TABLE 83 – Fit indices for SEM multi-group model by gender

	CFI	TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.901	0.899	0.949	0.073	0.116	107856541	2182522987	2181930798
Model 2	0.901	0.899	0.95	0.073	0.116	107546245	2182212691	2181620502
Model 3	0.901	0.899	0.949	0.074	0.116	107929487	2182595933	2182003744
Model 4	0.901	0.898	0.95	0.073	0.117	107539941	2182228892	2181618198
Model 5	0.901	0.898	0.949	0.073	0.117	107854734	2182543686	2181932991
Model 6	0.901	0.898	0.95	0.073	0.117	107538855	2182227807	2181617112
Model 7	0.901	0.897	0.95	0.073	0.117	107535483	2182246941	2181617740

TABLE 84 – Fit indices for multi-group estimates by area

	CFI	TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.924	0.909	0.959	0.061	0.109	80318830	2213068717	2212180434
Model 2	0.925	0.909	0.96	0.059	0.109	79792398	2212542284	2211654001
Model 3	0.924	0.909	0.959	0.061	0.109	80347110	2213096997	2212208714
Model 4	0.925	0.908	0.96	0.059	0.11	79786421	2212558813	2211652024
Model 5	0.925	0.908	0.959	0.061	0.11	80299387	2213071780	2212164991
Model 6	0.925	0.908	0.96	0.059	0.11	79787464	2212559856	2211653067
Model 7	0.925	0.907	0.96	0.061	0.11	79782900	2212577799	2211652504

TABLE 85 – Fit indices for multi-group estimates by school

	CFI	TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.913	0.913	0.955	0.064	0.107	92419352	2204215094	2203641411
Model 2	0.914	0.913	0.955	0.064	0.107	92136060	2203931802	2203358119
Model 3	0.914	0.913	0.955	0.065	0.107	92444508	2204240250	2203666568
Model 4	0.914	0.912	0.955	0.064	0.108	92129455	2203947703	2203355514
Model 5	0.913	0.912	0.955	0.064	0.108	92407797	2204226045	2203633857
Model 6	0.914	0.912	0.955	0.064	0.108	92127397	2203945646	2203353457
Model 7	0.914	0.911	0.955	0.064	0.108	92124905	2203965659	2203354965

TABLE 86 – Fit indices for multi-group estimates by color/race

	CFI	TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.92	0.904	0.957	0.06	0.113	85486090	2210329559	2208108852
Model 2	0.921	0.904	0.958	0.059	0.112	85028596	2209872065	2207651357
Model 3	0.92	0.904	0.957	0.061	0.113	85530954	2210374424	2208153716
Model 4	0.92	0.903	0.958	0.059	0.113	85018503	2209918238	2207651265
Model 5	0.92	0.903	0.957	0.06	0.113	85471900	2210371634	2208104662
Model 6	0.921	0.903	0.958	0.059	0.113	85018832	2209918566	2207651593
Model 7	0.921	0.902	0.958	0.059	0.114	85007633	2209963632	2207650395

TABLE 87 – Fit indices for multi-group estimates by region

	CFI	TLI	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Model 1	0.909	0.916	0.951	0.068	0.105	97159485	2199704593	2198705275
Model 2	0.91	0.916	0.952	0.067	0.105	96641041	2199186149	2198186830
Model 3	0.909	0.916	0.951	0.068	0.105	97225446	2199770553	2198771235
Model 4	0.91	0.916	0.952	0.067	0.105	96639239	2199240612	2198195028
Model 5	0.909	0.915	0.951	0.068	0.106	97144167	2199745539	2198699956
Model 6	0.91	0.916	0.952	0.067	0.105	96635491	2199236864	2198191281
Model 7	0.91	0.915	0.952	0.067	0.106	96631201	2199288838	2198196990

TABLE 88 – Fit Indices for mediation models - Part 1

		CFI Robust	TLI Robust	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Bullying	Alcool	0.925	0.913	0.902	0.048	0.078	86086218	2956015820	2956469214
	Drug	0.924	0.913	0.902	0.051	0.079	86697923	2957080919	2956627525
	Rsex	0.925	0.913	0.903	0.050	0.078	86067840	2956450836	2955997442
	g	0.925	0.913	0.903	0.048	0.078	85807381	2956190377	2955736983
Bully victimizer	Alcool	0.923	0.911	0.903	0.048	0.076	88964517	3024775184	3024294031
	Drug	0.921	0.909	0.901	0.054	0.077	91463199	3027273866	3026792713
	Rsex	0.921	0.910	0.902	0.053	0.076	90706486	3026517153	3026036000
	g	0.924	0.912	0.905	0.046	0.075	87852355	3023663022	3023181868
Peer effect	Alcool	0.910	0.922	0.899	0.056	0.080	89658715	2960041712	2959588317
	Drug	0.913	0.899	0.889	0.079	0.084	99759750	2970142746	2969689352
	Rsex	0.906	0.892	0.884	0.091	0.087	106962240	2977345236	2976891842
	g	0.925	0.913	0.903	0.047	0.078	85878225	2956261221	2955807827
Age	Alcool	0.916	0.904	0.896	0.055	0.079	97254926	3175940385	3175459231
	Drug	0.914	0.901	0.893	0.063	0.080	99854057	3178539515	3178058362
	Rsex	0.920	0.908	0.900	0.048	0.077	92623412	3171308870	3170827717
	g	0.918	0.905	0.897	0.051	0.079	95722838	3174408297	3173927143
N household	Alcool	0.925	0.914	0.906	0.046	0.074	86217583	3232929187	3232448034
	Drug	0.925	0.913	0.906	0.046	0.075	86317156	3233028760	3232547607
	Rsex	0.925	0.914	0.907	0.045	0.074	85719916	3232431520	3231950367
	g	0.925	0.914	0.906	0.046	0.074	86160012	3232871615	3232390462
Internet	Alcool	0.924	0.913	0.905	0.047	0.075	87110474	3034995055	3034513902
	Drug	0.924	0.913	0.905	0.047	0.075	87364340	3035248922	3034767768
	Rsex	0.924	0.913	0.906	0.047	0.075	86758773	3034643355	3034162201
	g	0.924	0.913	0.905	0.047	0.075	87120753	3035005335	3034524182

TABLE 89 – Fit Indices for mediation models - Part 2

		CFI Robust	TLI Robust	GFI	RMSR	RMSEA	Chi-square	BIC	AIC
Parents Smoke	Alcool	0.925	0.913	0.906	0.046	0.075	86374134	3040865499	3040384346
	Drug	0.924	0.912	0.905	0.049	0.075	87385843	3041877208	3041396055
	Rsex	0.924	0.913	0.906	0.047	0.075	86605251	3041096617	3040615463
	g	0.925	0.914	0.906	0.045	0.074	86018057	3040509423	3040028270
Escape classes	Alcool	0.924	0.913	0.905	0.048	0.075	87432026	3032170829	3031689676
	Drug	0.923	0.911	0.904	0.052	0.076	88709035	3033447838	3032966685
	Rsex	0.924	0.913	0.905	0.049	0.075	87512787	3032251591	3031770437
	g	0.925	0.914	0.906	0.045	0.075	86263018	3031001821	3030520668
HPV vaccine	Alcool	0.920	0.908	0.900	0.053	0.077	92005007	3062750102	3062268949
	Drug	0.920	0.908	0.901	0.053	0.077	91992913	3062738007	3062256854
	Rsex	0.924	0.912	0.905	0.046	0.075	87893229	3058638324	3058157170
	g	0.920	0.909	0.901	0.051	0.077	91619616	3062364711	3061883558
Comprehen-sive parents	Alcool	0.922	0.910	0.903	0.048	0.076	89997035	3065044082	3064562929
	Drug	0.920	0.908	0.901	0.052	0.077	91415506	3066462553	3065981400
	Rsex	0.921	0.909	0.902	0.051	0.077	90920297	3065967344	3065486191
	g	0.922	0.910	0.903	0.048	0.076	89693429	3064740477	3064259323
Housemaid	Alcool	0.925	0.913	0.906	0.046	0.075	86410006	3002561023	3002079870
	Drug	0.924	0.913	0.906	0.046	0.075	86567188	3002718205	3002237052
	Rsex	0.925	0.914	0.907	0.045	0.075	86005053	3002156071	3001674917
	g	0.925	0.913	0.906	0.046	0.075	86356893	3002507910	3002026757
Car	Alcool	0.924	0.913	0.905	0.047	0.075	86986226	3065577453	3065096299
	Drug	0.924	0.913	0.905	0.047	0.075	87037225	3065680557	3065199404
	Rsex	0.925	0.914	0.906	0.046	0.075	86264841	3064856068	3064374914
	g	0.924	0.913	0.906	0.047	0.075	86946382	3065537608	3065056455

APÊNDICE C – 3

TABLE 90 – PSM bullying

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.5221	0.4305	0.1356	0.0916	0.4499	0.1285	0.0722	21.2241
north	0.3425	0.3695	0.4827	-0.0269	0.3569	0.4791	-0.0143	46.8397
mid	0.1452	0.1322	0.3388	0.0130	0.1370	0.3439	0.0082	36.9656
southe	0.1789	0.1723	0.3776	0.0066	0.1756	0.3805	0.0033	49.9632
south	0.0990	0.0931	0.2906	0.0059	0.0958	0.2943	0.0032	45.5144
capital	0.5015	0.5054	0.5000	-0.0038	0.5026	0.5000	-0.0011	71.4461
sex	0.5517	0.5003	0.5000	0.0514	0.5057	0.5000	0.0460	10.6495
white	0.3232	0.3402	0.4738	-0.0170	0.3316	0.4708	-0.0084	50.6035
indigenous	0.0379	0.0367	0.1881	0.0011	0.0367	0.1879	0.0012	-5.6957
yellow	0.0467	0.0436	0.2042	0.0031	0.0445	0.2061	0.0023	27.6728
brown	0.4623	0.4603	0.4984	0.0021	0.4625	0.4986	-0.0002	92.4016
age	14.2931	14.3449	1.0732	-0.0518	14.3090	1.0434	-0.0159	69.2234
livesmother	0.8792	0.8942	0.3076	-0.0150	0.8914	0.3112	-0.0122	18.7702
livesfather	0.5967	0.6481	0.4776	-0.0514	0.6321	0.4822	-0.0354	31.1045
nhousehold	4.4869	4.5530	1.5951	-0.0660	4.5298	1.5930	-0.0428	35.1114
cellphone	0.8742	0.8761	0.3295	-0.0019	0.8754	0.3303	-0.0012	35.4770
internet	0.7727	0.7748	0.4177	-0.0021	0.7768	0.4164	-0.0042	-94.6556
car	0.5310	0.5489	0.4976	-0.0179	0.5434	0.4981	-0.0124	30.9896
houkee	0.1099	0.1262	0.3320	-0.0163	0.1171	0.3216	-0.0073	55.3268
works	0.1339	0.1148	0.3188	0.0191	0.1209	0.3261	0.0130	31.9609
hun30	1.4309	1.3058	0.7265	0.1251	1.3305	0.7512	0.1004	19.7795
maxschoo	0.6124	0.5622	0.4961	0.0502	0.5860	0.4926	0.0264	47.4577
ordstpreg	0.9100	0.8934	0.3086	0.0167	0.9061	0.2916	0.0039	76.6786
urban	0.9295	0.9137	0.2808	0.0158	0.9214	0.2690	0.0081	48.9690
public	0.7981	0.7835	0.4119	0.0146	0.7902	0.4071	0.0079	46.2689
missed30	0.0230	0.0191	0.1369	0.0039	0.0193	0.1376	0.0037	5.2632
mealwparents	2.2217	1.9478	1.6761	0.2740	2.0019	1.7100	0.2199	19.7424
parentsmoke	0.2570	0.2209	0.4149	0.0360	0.2311	0.4216	0.0258	28.3197
comparents30	3.0058	3.3437	1.4298	-0.3379	3.2633	1.4267	-0.2575	23.7858
attackfam30	1.4451	1.2483	0.9798	0.1967	1.2718	1.0256	0.1733	11.9173
sexviol	0.0543	0.0250	0.1561	0.0293	0.0277	0.1640	0.0266	9.0868
hourstv	3.7936	3.5574	2.6451	0.2362	3.6348	2.6624	0.1588	32.7628
feelalone12	2.6666	2.0646	1.1205	0.6020	2.1777	1.1229	0.4889	18.7902
hpv	0.3947	0.3629	0.4808	0.0318	0.3671	0.4820	0.0276	13.2642
healthserv12	0.5781	0.5433	0.4981	0.0348	0.5611	0.4963	0.0170	51.2379
thin	0.2687	0.2501	0.4331	0.0186	0.2657	0.4417	0.0030	83.9874
fat	0.2379	0.1297	0.3359	0.1082	0.1439	0.3510	0.0940	13.1569

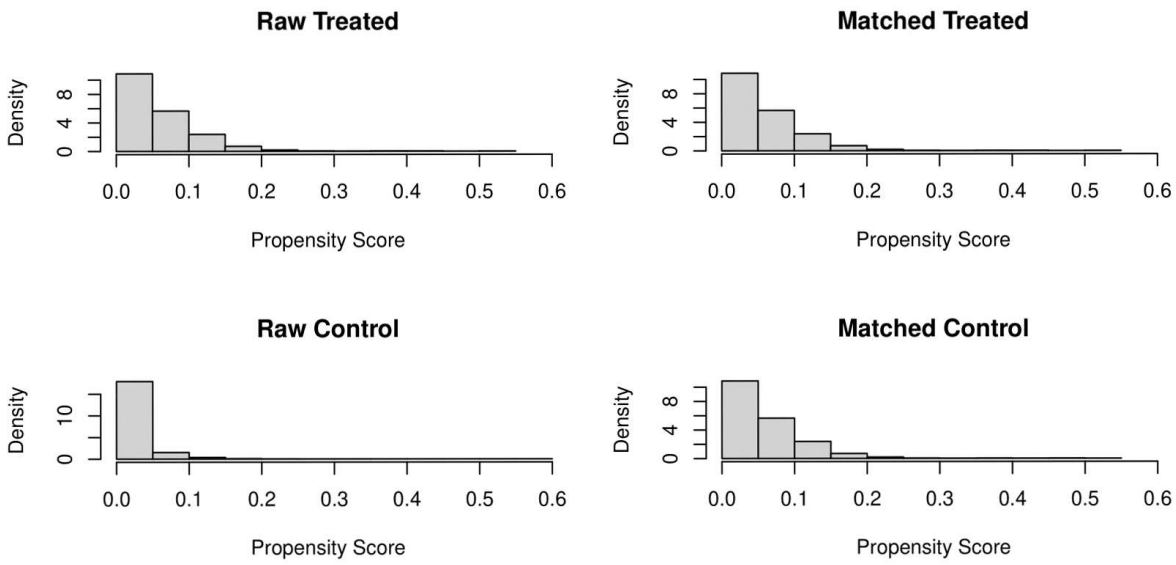

FIGURE 13 – PSM by motive - Color

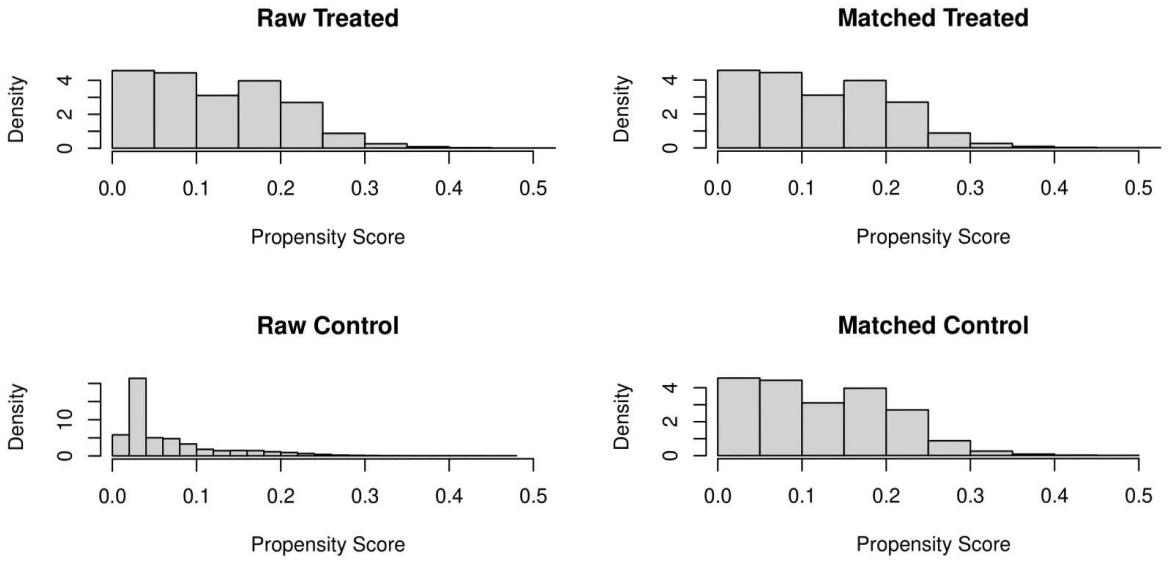

FIGURE 14 – PSM by motive - Body appearance

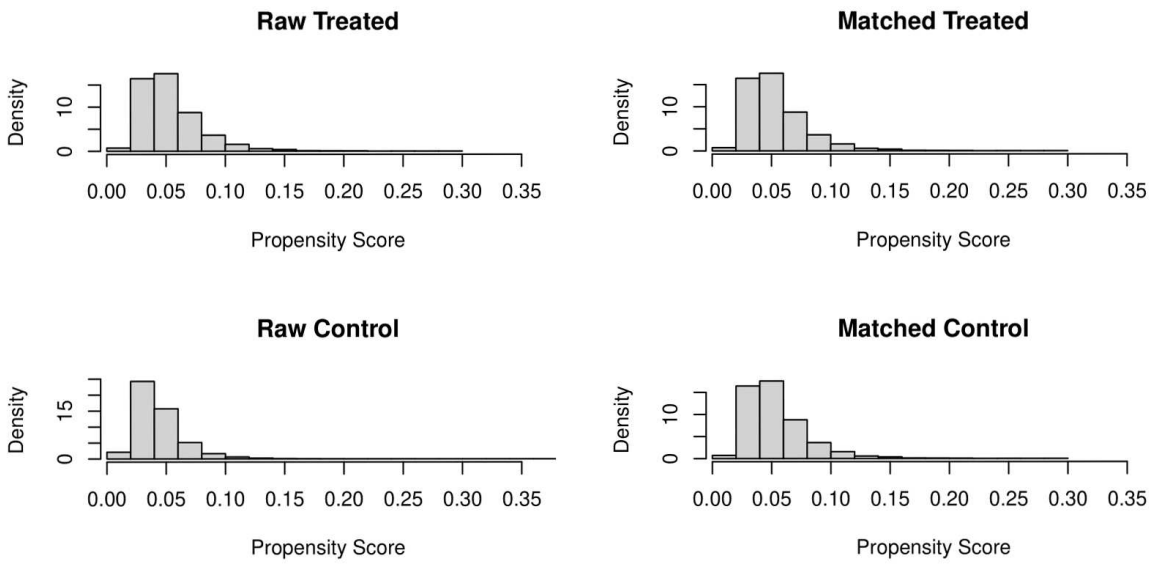

FIGURE 15 – PSM by motive - Facial Appearance

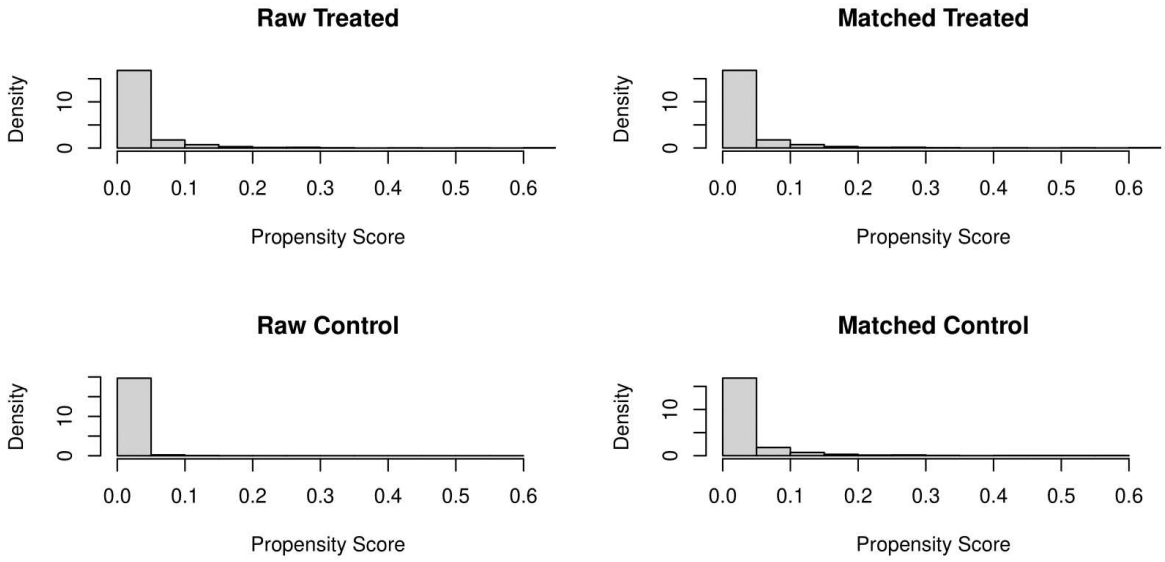

FIGURE 16 – PSM by motive - Sexual Orientation

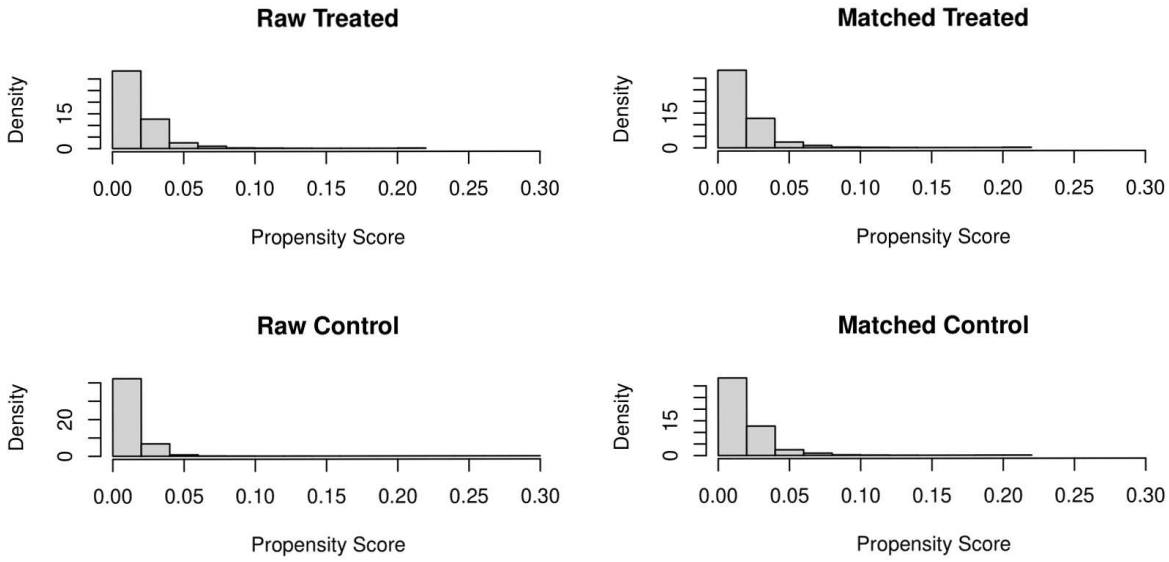

FIGURE 17 – PSM by motive - Religion

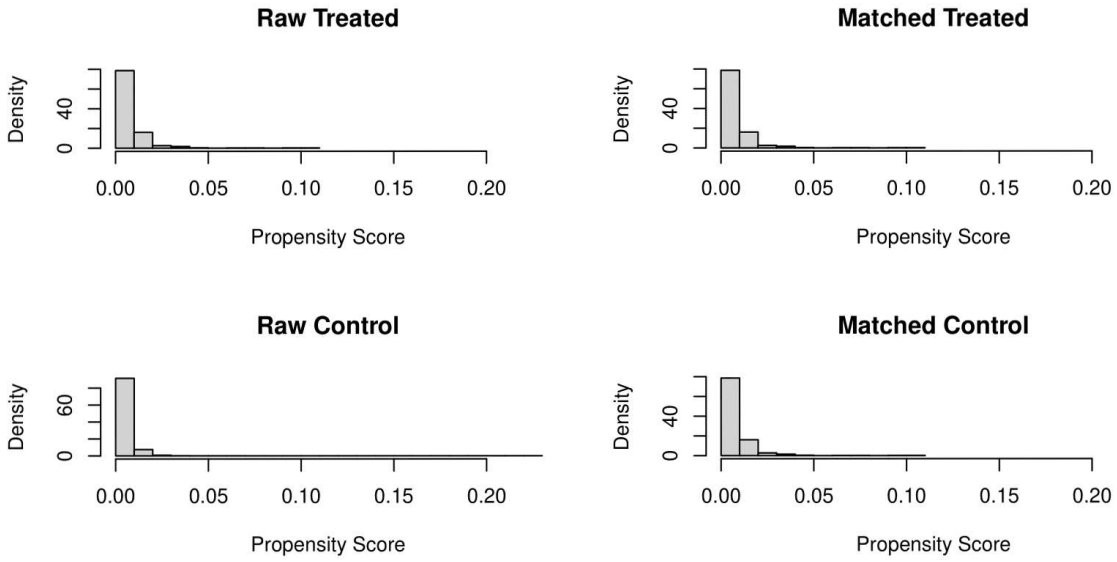

FIGURE 18 – PSM by motive - Region

TABLE 91 – PSM victimized

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.1120	0.0627	0.0507	0.0493	0.1118	0.0932	0.0001	99.7036
north	0.3199	0.3593	0.4798	-0.0393	0.3118	0.4632	0.0082	79.1825
mid	0.1624	0.1367	0.3435	0.0257	0.1720	0.3774	-0.0096	62.4576
southe	0.1929	0.1742	0.3793	0.0187	0.1873	0.3902	0.0056	69.9783
south	0.0891	0.0964	0.2951	-0.0072	0.0895	0.2854	-0.0003	95.5649
capital	0.4912	0.5044	0.5000	-0.0132	0.4910	0.5000	0.0002	98.7871
sex	0.4957	0.5267	0.4993	-0.0310	0.4828	0.4997	0.0128	58.5802
white	0.3121	0.3336	0.4715	-0.0215	0.3108	0.4629	0.0013	94.0226
indigenous	0.0435	0.0368	0.1883	0.0067	0.0442	0.2055	-0.0006	90.4271
yellow	0.0483	0.0449	0.2070	0.0035	0.0488	0.2155	-0.0005	86.1714
brown	0.4537	0.4618	0.4985	-0.0080	0.4639	0.4987	-0.0101	-25.7314
age	14.2933	14.3223	1.0472	-0.0290	14.2727	0.9974	0.0206	29.0306
livesmother	0.8799	0.8876	0.3159	-0.0077	0.8816	0.3231	-0.0018	77.1204
livesfather	0.5948	0.6258	0.4839	-0.0310	0.6009	0.4898	-0.0061	80.3283
nhousehold	4.4746	4.5250	1.5934	-0.0504	4.4687	1.6125	0.0059	88.2001
cellphone	0.8526	0.8768	0.3287	-0.0242	0.8519	0.3552	0.0006	97.3458
internet	0.7417	0.7761	0.4169	-0.0344	0.7448	0.4360	-0.0031	91.1161
car	0.5218	0.5417	0.4983	-0.0199	0.5172	0.4997	0.0047	76.5682
houkee	0.1217	0.1182	0.3229	0.0035	0.1136	0.3173	0.0082	-131.5358
work	0.1656	0.1210	0.3261	0.0446	0.1614	0.3679	0.0042	90.6459
hun30	1.5891	1.3493	0.7583	0.2399	1.5737	0.9635	0.0154	93.5722
maxschoo	0.5686	0.5872	0.4923	-0.0186	0.5781	0.4939	-0.0095	49.1158
ordstpreg	0.8818	0.9026	0.2964	-0.0209	0.8783	0.3270	0.0035	83.0558
urban	0.9261	0.9208	0.2700	0.0053	0.9243	0.2645	0.0018	66.4519
public	0.8087	0.7891	0.4079	0.0196	0.8143	0.3889	-0.0056	71.2944
missed30	0.0411	0.0195	0.1383	0.0216	0.0411	0.1986	0.0000	100.0000
mealwparents	2.3925	2.0554	1.7384	0.3371	2.3675	1.9045	0.0251	92.5682
parentsmoke	0.2644	0.2362	0.4247	0.0282	0.2626	0.4401	0.0018	93.7385
comparents30	2.8681	3.2058	1.4393	-0.3377	2.8680	1.4495	0.0002	99.9524
attackfam30	1.8371	1.3066	1.0490	0.5305	1.7888	1.7694	0.0483	90.8873
sexviol	0.0901	0.0353	0.1844	0.0548	0.0893	0.2852	0.0008	98.5357
hourstv	4.0654	3.6414	2.6790	0.4240	4.0991	2.8859	-0.0337	92.0447
feelalone12	3.0092	2.3033	1.1928	0.7058	3.0120	1.3101	-0.0029	99.5904
hpv	0.3620	0.3791	0.4852	-0.0171	0.3574	0.4793	0.0047	72.7204
healthserv12	0.5700	0.5590	0.4965	0.0110	0.5681	0.4954	0.0019	82.4475
thin	0.2922	0.2566	0.4367	0.0356	0.2939	0.4556	-0.0018	95.0365
fat	0.2820	0.1738	0.3790	0.1082	0.2812	0.4496	0.0008	99.2579

TABLE 92 – PSM victmizer

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.2357	0.1735	0.0888	0.0622	0.2348	0.1157	0.0009	98.4910
north	0.3222	0.3645	0.4813	-0.0424	0.3199	0.4665	0.0022	94.7370
mid	0.1494	0.1359	0.3427	0.0134	0.1528	0.3598	-0.0034	74.4520
southe	0.1955	0.1708	0.3764	0.0247	0.1973	0.3980	-0.0018	92.5805
south	0.0964	0.0958	0.2943	0.0006	0.0945	0.2925	0.0019	-222.5768
capital	0.5243	0.4988	0.5000	0.0255	0.5308	0.4991	-0.0065	74.4201
sex	0.4067	0.5514	0.4974	-0.1447	0.4032	0.4906	0.0035	97.5889
white	0.3162	0.3358	0.4723	-0.0196	0.3151	0.4646	0.0011	94.4460
indigenous	0.0405	0.0365	0.1876	0.0040	0.0408	0.1979	-0.0003	91.3690
yellow	0.0502	0.0439	0.2049	0.0063	0.0501	0.2182	0.0001	98.1834
brown	0.4548	0.4627	0.4986	-0.0079	0.4553	0.4980	-0.0005	93.4946
age	14.3777	14.3073	1.0503	0.0704	14.3764	1.0663	0.0013	98.1307
livesmother	0.8734	0.8902	0.3127	-0.0168	0.8739	0.3320	-0.0005	97.2697
livesfather	0.5903	0.6313	0.4824	-0.0411	0.5909	0.4917	-0.0006	98.4684
nhousehold	4.4949	4.5278	1.5890	-0.0329	4.5018	1.6096	-0.0069	78.9646
cellphone	0.8831	0.8734	0.3326	0.0097	0.8816	0.3231	0.0015	84.6881
internet	0.8021	0.7674	0.4225	0.0348	0.8035	0.3974	-0.0013	96.2145
car	0.5605	0.5359	0.4987	0.0246	0.5611	0.4963	-0.0006	97.4424
houkee	0.1472	0.1119	0.3153	0.0353	0.1475	0.3546	-0.0002	99.3524
works	0.1757	0.1121	0.3155	0.0635	0.1753	0.3803	0.0003	99.4597
hun30	1.4677	1.3418	0.7527	0.1259	1.4712	0.8872	-0.0035	97.2288
maxschoo	0.5702	0.5896	0.4919	-0.0193	0.5762	0.4942	-0.0059	69.1903
ordstpreg	0.8957	0.9025	0.2966	-0.0068	0.8960	0.3053	-0.0003	95.7962
urban	0.9329	0.9185	0.2735	0.0144	0.9334	0.2493	-0.0005	96.4158
public	0.7796	0.7929	0.4053	-0.0133	0.7778	0.4158	0.0018	86.2010
missed30	0.0411	0.0164	0.1268	0.0248	0.0342	0.1818	0.0069	72.0642
mealwparents	2.3331	2.0196	1.7190	0.3135	2.3288	1.8840	0.0043	98.6135
parentsmoke	0.2742	0.2298	0.4207	0.0444	0.2768	0.4474	-0.0026	94.0680
comparents30	2.8890	3.2504	1.4459	-0.3615	2.8677	1.4532	0.0213	94.1129
attackfam30	1.6372	1.2744	0.9974	0.3627	1.5970	1.5592	0.0402	88.9293
sexviol	0.0720	0.0314	0.1743	0.0406	0.0675	0.2509	0.0045	89.0088
hourstv	4.1761	3.5543	2.6527	0.6218	4.1786	2.8514	-0.0025	99.5952
feelalone12	2.4875	2.3186	1.2133	0.1689	2.4923	1.2844	-0.0047	97.1898
hpv	0.2897	0.3980	0.4895	-0.1083	0.2868	0.4523	0.0030	97.2533
healthserv12	0.5734	0.5567	0.4968	0.0168	0.5761	0.4942	-0.0026	84.3192
thin	0.2826	0.2535	0.4350	0.0290	0.2819	0.4500	0.0006	97.8331
fat	0.1934	0.1781	0.3826	0.0153	0.1944	0.3957	-0.0010	93.6310

TABLE 93 – PSM by motive - Color

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.0586	0.0213	0.0283	0.0372	0.0585	0.0565	0.0000	99.9729
north	0.3228	0.3575	0.4793	-0.0347	0.3214	0.4671	0.0014	95.8605
mid	0.1683	0.1377	0.3446	0.0306	0.1597	0.3664	0.0086	71.8797
southe	0.1703	0.1755	0.3804	-0.0053	0.1664	0.3726	0.0038	27.4413
south	0.0832	0.0962	0.2949	-0.0130	0.0913	0.2882	-0.0081	37.3718
capital	0.4720	0.5042	0.5000	-0.0322	0.4778	0.4996	-0.0057	82.1915
sex	0.3797	0.5279	0.4992	-0.1482	0.3678	0.4823	0.0120	91.9324
white	0.1153	0.3371	0.4727	-0.2218	0.1038	0.3050	0.0115	94.8255
indigenous	0.0727	0.0365	0.1874	0.0362	0.0803	0.2719	-0.0077	78.8835
yellow	0.0340	0.0453	0.2081	-0.0114	0.0306	0.1723	0.0033	70.6030
brown	0.3214	0.4644	0.4987	-0.1430	0.3367	0.4727	-0.0153	89.3018
age	14.5591	14.3149	1.0451	0.2441	14.5705	1.1897	-0.0115	95.2983
livesmother	0.8828	0.8872	0.3164	-0.0043	0.8895	0.3136	-0.0067	-54.3902
livesfather	0.6093	0.6241	0.4844	-0.0148	0.6184	0.4859	-0.0091	38.5849
nhousehold	4.7661	4.5161	1.5932	0.2500	4.7929	1.7802	-0.0268	89.2877
cellphone	0.8436	0.8759	0.3297	-0.0323	0.8355	0.3708	0.0081	74.8003
internet	0.6619	0.7763	0.4167	-0.1145	0.6490	0.4774	0.0129	88.7192
car	0.4548	0.5423	0.4982	-0.0875	0.4476	0.4974	0.0072	91.8057
houkee	0.1086	0.1187	0.3234	-0.0101	0.1019	0.3025	0.0067	33.7483
works	0.1717	0.1228	0.3282	0.0489	0.1640	0.3704	0.0077	84.3444
hun30	1.5782	1.3602	0.7728	0.2179	1.5595	0.9815	0.0187	91.4424
maxschoo	0.4840	0.5883	0.4921	-0.1043	0.4773	0.4996	0.0067	93.5812
ordstpreg	0.8651	0.9021	0.2972	-0.0369	0.8714	0.3349	-0.0062	83.1730
urban	0.8876	0.9220	0.2682	-0.0343	0.8895	0.3136	-0.0019	94.4299
public	0.8431	0.7892	0.4079	0.0539	0.8551	0.3521	-0.0120	77.8257
missed30	0.0330	0.0207	0.1423	0.0123	0.0330	0.1787	0.0000	100.0000
mealwparents	2.0928	2.0773	1.7539	0.0155	2.0746	1.7814	0.0182	-17.2148
parentsmoke	0.2927	0.2368	0.4251	0.0559	0.2879	0.4529	0.0048	91.4450
comparents30	3.0626	3.1863	1.4477	-0.1236	3.0545	1.4644	0.0081	93.4236
attackfam30	1.6872	1.3337	1.1055	0.3535	1.6074	1.6033	0.0799	77.4062
sexviol	0.0875	0.0378	0.1907	0.0497	0.0813	0.2734	0.0062	87.5011
hourstv	3.8737	3.6647	2.6998	0.2090	3.9570	2.8477	-0.0832	60.1939
feelalone12	2.4825	2.3468	1.2178	0.1357	2.5041	1.2883	-0.0215	84.1412
hpv	0.2630	0.3806	0.4855	-0.1175	0.2563	0.4367	0.0067	94.3042
healthserv12	0.5610	0.5597	0.4964	0.0012	0.5509	0.4975	0.0100	-715.8028
thin	0.2865	0.2583	0.4377	0.0282	0.2903	0.4540	-0.0038	86.4219
fat	0.1406	0.1819	0.3857	-0.0413	0.1339	0.3406	0.0067	83.7754

TABLE 94 – PSM by motive - Body appearance

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.1254	0.0642	0.0595	0.0613	0.1254	0.0790	0.0000	99.9782
north	0.3460	0.3575	0.4793	-0.0115	0.3516	0.4775	-0.0056	51.3602
mid	0.1562	0.1371	0.3440	0.0191	0.1571	0.3640	-0.0009	95.1409
southe	0.1914	0.1742	0.3793	0.0171	0.1892	0.3917	0.0022	87.3406
south	0.1008	0.0955	0.2940	0.0052	0.0997	0.2996	0.0011	79.3171
capital	0.5140	0.5028	0.5000	0.0112	0.5118	0.4999	0.0022	80.7245
sex	0.5707	0.5213	0.4996	0.0494	0.5766	0.4941	-0.0059	88.0938
white	0.3637	0.3298	0.4702	0.0338	0.3627	0.4808	0.0009	97.2544
indigenous	0.0389	0.0371	0.1891	0.0017	0.0387	0.1929	0.0002	90.9984
yellow	0.0447	0.0451	0.2076	-0.0004	0.0451	0.2074	-0.0003	16.8773
brown	0.4640	0.4611	0.4985	0.0029	0.4674	0.4990	-0.0034	-15.6548
age	14.1804	14.3306	1.0540	-0.1502	14.1770	0.9654	0.0034	97.7330
livesmother	0.8878	0.8870	0.3166	0.0007	0.8788	0.3264	0.0090	-1127.2456
livesfather	0.6012	0.6254	0.4840	-0.0242	0.5998	0.4900	0.0014	94.2481
nhousehold	4.3764	4.5323	1.6019	-0.1560	4.3925	1.5599	-0.0161	89.6755
cellphone	0.8817	0.8747	0.3311	0.0070	0.8783	0.3270	0.0034	51.5638
internet	0.7961	0.7722	0.4194	0.0239	0.7879	0.4088	0.0082	65.6929
car	0.5832	0.5373	0.4986	0.0459	0.5868	0.4925	-0.0036	92.2498
houkee	0.1265	0.1179	0.3224	0.0086	0.1265	0.3324	0.0000	100.0000
works	0.1266	0.1237	0.3292	0.0030	0.1240	0.3296	0.0026	10.8701
hun30	1.4527	1.3586	0.7743	0.0941	1.4333	0.8467	0.0194	79.4264
maxschoo	0.6359	0.5823	0.4932	0.0535	0.6340	0.4817	0.0019	96.5298
ordstpreg	0.9076	0.9008	0.2989	0.0068	0.9118	0.2837	-0.0042	38.2318
urban	0.9390	0.9199	0.2715	0.0191	0.9367	0.2436	0.0023	87.8471
public	0.7695	0.7919	0.4059	-0.0225	0.7707	0.4204	-0.0012	94.4904
missed30	0.0217	0.0209	0.1430	0.0008	0.0231	0.1501	-0.0014	-75.7458
mealwparents	2.2534	2.0647	1.7480	0.1887	2.2143	1.8288	0.0392	79.2454
parentsmoke	0.2677	0.2358	0.4245	0.0318	0.2618	0.4397	0.0059	81.5269
comparents30	2.9779	3.1986	1.4475	-0.2208	2.9889	1.4696	-0.0110	95.0209
attackfam30	1.5228	1.3283	1.1009	0.1946	1.5194	1.4130	0.0034	98.2494
sexviol	0.0553	0.0377	0.1904	0.0176	0.0560	0.2300	-0.0008	95.6009
hourstv	3.8607	3.6553	2.7013	0.2054	3.8413	2.7645	0.0194	90.5769
feelalone12	2.7545	2.3202	1.2108	0.4343	2.7470	1.2822	0.0074	98.2888
hpv	0.4191	0.3750	0.4841	0.0441	0.4236	0.4942	-0.0045	89.8289
healthserv12	0.6033	0.5566	0.4968	0.0468	0.6049	0.4889	-0.0015	96.6897
thin	0.2951	0.2563	0.4366	0.0388	0.2968	0.4569	-0.0017	95.6146
fat	0.4810	0.1589	0.3656	0.3221	0.4812	0.4997	-0.0002	99.9519

TABLE 95 – PSM by motive - Facial Appearance

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.0538	0.0436	0.0209	0.0103	0.0538	0.0269	0.0000	99.9952
north	0.3470	0.3571	0.4792	-0.0101	0.3519	0.4776	-0.0048	52.4108
mid	0.1251	0.1390	0.3460	-0.0140	0.1195	0.3244	0.0055	60.3754
southe	0.2215	0.1733	0.3785	0.0482	0.2253	0.4179	-0.0038	92.0187
south	0.0890	0.0962	0.2949	-0.0072	0.0851	0.2791	0.0038	46.8623
capital	0.5031	0.5036	0.5000	-0.0004	0.4848	0.4998	0.0183	-4174.5031
sex	0.4608	0.5276	0.4992	-0.0668	0.4625	0.4987	-0.0017	97.4791
white	0.3122	0.3331	0.4713	-0.0209	0.2896	0.4536	0.0226	-8.1191
indigenous	0.0373	0.0373	0.1894	0.0000	0.0370	0.1889	0.0002	-1085.2756
yellow	0.0539	0.0447	0.2066	0.0092	0.0570	0.2319	-0.0031	65.9626
brown	0.4764	0.4606	0.4984	0.0159	0.4925	0.5000	-0.0161	-1.5382
age	14.2838	14.3220	1.0487	-0.0382	14.2874	1.0094	-0.0036	90.5658
livesmother	0.8870	0.8871	0.3165	-0.0001	0.8877	0.3158	-0.0007	-544.8522
livesfather	0.6224	0.6238	0.4844	-0.0014	0.6130	0.4871	0.0094	-573.6360
nhousehold	4.4940	4.5229	1.5970	-0.0290	4.5038	1.5917	-0.0099	65.9438
cellphone	0.8545	0.8761	0.3295	-0.0216	0.8405	0.3661	0.0139	35.4754
internet	0.7605	0.7744	0.4180	-0.0140	0.7588	0.4279	0.0017	87.9479
car	0.5257	0.5411	0.4983	-0.0154	0.5291	0.4992	-0.0034	78.0717
houkee	0.1089	0.1189	0.3236	-0.0099	0.1051	0.3067	0.0038	61.2618
works	0.1354	0.1234	0.3289	0.0120	0.1316	0.3380	0.0038	68.0333
hun30	1.5065	1.3586	0.7727	0.1479	1.4916	0.8981	0.0149	89.9208
maxschoo	0.5688	0.5868	0.4924	-0.0180	0.5503	0.4975	0.0185	-2.9486
ordstpreg	0.9016	0.9012	0.2983	0.0004	0.9045	0.2939	-0.0029	-646.7383
urban	0.9382	0.9204	0.2707	0.0178	0.9365	0.2439	0.0017	90.5298
public	0.7953	0.7902	0.4072	0.0051	0.7953	0.4035	0.0000	100.0000
missed30	0.0269	0.0207	0.1422	0.0063	0.0301	0.1708	-0.0031	50.1808
mealwparents	2.1934	2.0723	1.7514	0.1211	2.2398	1.8435	-0.0464	61.6624
parentsmoke	0.2667	0.2367	0.4251	0.0300	0.2619	0.4397	0.0048	83.9761
comparents30	2.9935	3.1923	1.4466	-0.1988	2.9767	1.4550	0.0168	91.5309
attackfam30	1.5659	1.3312	1.1041	0.2347	1.5633	1.4940	0.0026	98.8726
sexviol	0.0676	0.0376	0.1901	0.0300	0.0664	0.2490	0.0012	95.9949
hourstv	3.8401	3.6615	2.7018	0.1786	3.8514	2.7711	-0.0113	93.6712
feelalone12	2.6657	2.3353	1.2155	0.3304	2.6433	1.2730	0.0224	93.2305
hpv	0.3451	0.3795	0.4853	-0.0344	0.3458	0.4757	-0.0007	97.9011
healthserv12	0.5582	0.5598	0.4964	-0.0016	0.5447	0.4981	0.0135	-719.6968
thin	0.3124	0.2564	0.4367	0.0560	0.3175	0.4655	-0.0051	90.9749
fat	0.1643	0.1817	0.3856	-0.0175	0.1539	0.3609	0.0103	40.7801

TABLE 96 – PSM by motive - Sexual Orientation

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.0331	0.0083	0.0146	0.0248	0.0329	0.0562	0.0002	99.3127
north	0.3271	0.3570	0.4791	-0.0299	0.3695	0.4830	-0.0424	-42.1226
mid	0.1386	0.1384	0.3453	0.0002	0.1311	0.3377	0.0075	-4236.8056
southe	0.1973	0.1752	0.3802	0.0220	0.1860	0.3894	0.0112	48.9984
south	0.0886	0.0960	0.2946	-0.0073	0.0811	0.2732	0.0075	-2.1788
capital	0.5144	0.5034	0.5000	0.0109	0.5281	0.4995	-0.0137	-25.8242
sex	0.3546	0.5261	0.4993	-0.1715	0.3271	0.4694	0.0275	83.9886
white	0.3396	0.3321	0.4710	0.0075	0.3296	0.4704	0.0100	-33.5136
indigenous	0.0375	0.0373	0.1894	0.0002	0.0549	0.2280	-0.0175	-8785.1626
yellow	0.0512	0.0450	0.2074	0.0061	0.0612	0.2398	-0.0100	-62.4694
brown	0.4532	0.4613	0.4985	-0.0081	0.4482	0.4976	0.0050	38.6934
age	14.5830	14.3181	1.0472	0.2649	14.5968	1.2403	-0.0137	94.8163
livesmother	0.8365	0.8875	0.3160	-0.0511	0.8327	0.3735	0.0037	92.6635
livesfather	0.4956	0.6248	0.4842	-0.1292	0.4931	0.5003	0.0025	98.0676
nhousehold	4.4657	4.5221	1.5959	-0.0565	4.4132	1.7536	0.0524	7.1592
cellphone	0.9051	0.8749	0.3308	0.0302	0.8889	0.3145	0.0162	46.2782
internet	0.7928	0.7737	0.4185	0.0191	0.8002	0.4001	-0.0075	60.7932
car	0.5218	0.5406	0.4984	-0.0187	0.5156	0.5001	0.0062	66.6610
houkee	0.1760	0.1180	0.3226	0.0581	0.1536	0.3607	0.0225	61.3084
works	0.1898	0.1233	0.3288	0.0664	0.1898	0.3924	0.0000	100.0000
hun30	1.6429	1.3627	0.7757	0.2803	1.6467	1.0554	-0.0037	98.6636
maxschoo	0.6230	0.5857	0.4926	0.0373	0.6317	0.4826	-0.0087	76.5759
ordstpreg	0.8527	0.9017	0.2977	-0.0490	0.8627	0.3444	-0.0100	79.6162
urban	0.9413	0.9210	0.2697	0.0203	0.9438	0.2304	-0.0025	87.6992
public	0.8040	0.7903	0.4071	0.0137	0.7940	0.4047	0.0100	27.0919
missed30	0.0737	0.0205	0.1417	0.0532	0.0699	0.2552	0.0037	92.9562
mealwparents	2.5905	2.0732	1.7516	0.5173	2.5081	1.9667	0.0824	84.0711
parentsmoke	0.3258	0.2373	0.4254	0.0886	0.3171	0.4656	0.0087	90.1338
comparents30	2.6005	3.1885	1.4462	-0.5880	2.6130	1.4543	-0.0125	97.8769
attackfam30	2.1798	1.3344	1.1063	0.8454	2.1411	2.1104	0.0387	95.4220
sexviol	0.2135	0.0374	0.1897	0.1761	0.1960	0.3972	0.0175	90.0748
hourstv	4.2210	3.6646	2.6995	0.5564	4.1461	2.9364	0.0749	86.5365
feelalone12	3.1323	2.3432	1.2160	0.7892	3.1910	1.3341	-0.0587	92.5649
hpv	0.2397	0.3792	0.4852	-0.1395	0.2147	0.4109	0.0250	82.0963
healthserv12	0.5818	0.5596	0.4964	0.0222	0.5830	0.4934	-0.0012	94.3736
thin	0.2834	0.2587	0.4379	0.0247	0.2846	0.4515	-0.0012	94.9441
fat	0.2509	0.1804	0.3845	0.0706	0.2397	0.4272	0.0112	84.0802

TABLE 97 – PSM by motive - Religion

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.0201	0.0146	0.0090	0.0056	0.0201	0.0158	0.0000	99.9986
north	0.3883	0.3562	0.4789	0.0321	0.3912	0.4882	-0.0029	90.9761
mid	0.1374	0.1384	0.3453	-0.0010	0.1439	0.3511	-0.0065	-526.8045
southe	0.1359	0.1760	0.3808	-0.0401	0.1345	0.3413	0.0014	96.3901
south	0.0738	0.0962	0.2949	-0.0225	0.0622	0.2416	0.0116	48.5463
capital	0.4346	0.5046	0.5000	-0.0700	0.4534	0.4980	-0.0188	73.1423
sex	0.5235	0.5247	0.4994	-0.0012	0.5575	0.4969	-0.0340	-2833.4587
white	0.2972	0.3327	0.4712	-0.0355	0.2813	0.4498	0.0159	55.1878
indigenous	0.0506	0.0371	0.1889	0.0136	0.0528	0.2237	-0.0022	83.9969
yellow	0.0506	0.0450	0.2073	0.0056	0.0586	0.2349	-0.0080	-41.8825
brown	0.4729	0.4611	0.4985	0.0118	0.4772	0.4997	-0.0043	63.2269
age	14.5271	14.3173	1.0452	0.2098	14.4808	1.1599	0.0463	77.9470
livesmother	0.8749	0.8873	0.3163	-0.0123	0.8641	0.3428	0.0108	12.1278
livesfather	0.6298	0.6237	0.4845	0.0061	0.6233	0.4847	0.0065	-6.0837
nhousehold	4.7404	4.5184	1.5942	0.2220	4.8019	1.7482	-0.0615	72.3153
cellphone	0.8366	0.8757	0.3299	-0.0392	0.8344	0.3718	0.0022	94.4593
internet	0.6956	0.7750	0.4176	-0.0794	0.7050	0.4562	-0.0094	88.1597
car	0.4751	0.5414	0.4983	-0.0663	0.4772	0.4997	-0.0022	96.7296
houkee	0.1114	0.1185	0.3233	-0.0072	0.1157	0.3200	-0.0043	39.7109
works	0.1779	0.1231	0.3285	0.0548	0.1793	0.3838	-0.0014	97.3602
hun30	1.5539	1.3623	0.7755	0.1916	1.5539	0.9600	0.0000	100.0000
maxschoo	0.5409	0.5866	0.4924	-0.0458	0.5423	0.4984	-0.0014	96.8423
ordstpreg	0.8742	0.9017	0.2978	-0.0275	0.8698	0.3366	0.0043	84.2135
urban	0.8785	0.9218	0.2684	-0.0433	0.8800	0.3251	-0.0014	96.6607
public	0.8445	0.7896	0.4076	0.0549	0.8351	0.3712	0.0094	82.8884
missed30	0.0296	0.0208	0.1427	0.0088	0.0275	0.1635	0.0022	75.4575
mealwparents	2.1641	2.0763	1.7532	0.0878	2.2162	1.8249	-0.0521	40.7078
parentsmoke	0.2336	0.2381	0.4259	-0.0045	0.2278	0.4195	0.0058	-27.5742
comparents30	3.0860	3.1850	1.4474	-0.0989	3.1258	1.4977	-0.0398	59.8062
attackfam30	1.6905	1.3364	1.1102	0.3541	1.7129	1.7323	-0.0224	93.6706
sexviol	0.0918	0.0381	0.1914	0.0537	0.0904	0.2868	0.0014	97.3089
hourstv	3.3651	3.6738	2.7018	-0.3087	3.4591	2.7127	-0.0940	69.5496
feelalone12	2.5813	2.3464	1.2178	0.2349	2.6132	1.3101	-0.0318	86.4582
hpv	0.3384	0.3786	0.4850	-0.0402	0.3637	0.4812	-0.0253	37.0040
healthserv12	0.5864	0.5594	0.4965	0.0270	0.5936	0.4913	-0.0072	73.2498
thin	0.2610	0.2589	0.4380	0.0021	0.2574	0.4374	0.0036	-68.4801
fat	0.1829	0.1809	0.3850	0.0020	0.1844	0.3879	-0.0014	28.0190

TABLE 98 – PSM by motive - Region

Variables	Treated	All data			Paired data			% improvement
	Mean treat	Mean cont	SD cont	Diff mean	Mean cont	SD cont	Diff mean	Diff mean
distance	0.0086	0.0058	0.0041	0.0028	0.0086	0.0090	0.0000	99.6824
north	0.2980	0.3570	0.4791	-0.0591	0.3400	0.4742	-0.0420	28.8021
mid	0.1737	0.1382	0.3451	0.0355	0.1499	0.3573	0.0238	33.0053
southe	0.1682	0.1755	0.3804	-0.0073	0.1810	0.3854	-0.0128	-76.2308
south	0.1152	0.0958	0.2943	0.0194	0.1243	0.3302	-0.0091	52.8289
capital	0.5027	0.5035	0.5000	-0.0008	0.5229	0.4999	-0.0201	-2421.0302
sex	0.4113	0.5253	0.4994	-0.1140	0.4241	0.4947	-0.0128	88.7712
white	0.3309	0.3322	0.4710	-0.0013	0.3400	0.4742	-0.0091	-619.7514
indigenous	0.0548	0.0372	0.1891	0.0177	0.0585	0.2349	-0.0037	79.3299
yellow	0.0585	0.0450	0.2073	0.0135	0.0585	0.2349	0.0000	100.0000
brown	0.4442	0.4614	0.4985	-0.0171	0.4369	0.4965	0.0073	57.2806
age	14.4625	14.3195	1.0481	0.1430	14.4406	1.0968	0.0219	84.6594
livesmother	0.8665	0.8872	0.3164	-0.0206	0.8848	0.3195	-0.0183	11.4550
livesfather	0.6033	0.6239	0.4844	-0.0206	0.5704	0.4955	0.0329	-59.9422
nhousehold	4.5448	4.5215	1.5970	0.0233	4.4095	1.5325	0.1353	-481.6753
cellphone	0.8848	0.8751	0.3306	0.0097	0.8958	0.3058	-0.0110	-12.8660
internet	0.7660	0.7739	0.4183	-0.0079	0.7587	0.4283	0.0073	7.0206
car	0.5539	0.5403	0.4984	0.0136	0.5009	0.5005	0.0530	-289.9165
houkee	0.1408	0.1183	0.3230	0.0225	0.1280	0.3344	0.0128	43.0103
works	0.1883	0.1235	0.3290	0.0648	0.1901	0.3928	-0.0018	97.1779
hun30	1.5704	1.3639	0.7772	0.2065	1.4954	0.9166	0.0750	63.7049
maxschoo	0.5868	0.5860	0.4926	0.0009	0.5960	0.4912	-0.0091	-959.4532
ordstpreg	0.8757	0.9014	0.2981	-0.0257	0.8867	0.3173	-0.0110	57.3680
urban	0.9287	0.9212	0.2695	0.0075	0.9342	0.2482	-0.0055	27.3466
public	0.7733	0.7905	0.4069	-0.0172	0.8044	0.3970	-0.0311	-80.6578
missed30	0.0475	0.0208	0.1427	0.0268	0.0475	0.2130	0.0000	100.0000
mealwparents	2.2614	2.0765	1.7538	0.1849	2.3144	1.8840	-0.0530	71.3237
parentsmoke	0.2797	0.2378	0.4257	0.0419	0.2706	0.4447	0.0091	78.2011
comparents30	3.0329	3.1844	1.4475	-0.1515	2.8665	1.4458	0.1664	-9.8033
attackfam30	1.7038	1.3395	1.1157	0.3644	1.7989	1.8486	-0.0951	73.9108
sexviol	0.0841	0.0386	0.1927	0.0455	0.0932	0.2910	-0.0091	79.9019
hourstv	4.0238	3.6673	2.7040	0.3565	4.1974	2.7803	-0.1737	51.2843
feelalone12	2.6124	2.3483	1.2186	0.2641	2.6929	1.2626	-0.0804	69.5441
hpv	0.2870	0.3785	0.4850	-0.0915	0.2943	0.4562	-0.0073	92.0071
healthserv12	0.5740	0.5597	0.4964	0.0144	0.5795	0.4941	-0.0055	61.7848
thin	0.2486	0.2590	0.4381	-0.0103	0.2468	0.4315	0.0018	82.3253
fat	0.1682	0.1810	0.3850	-0.0128	0.1517	0.3591	0.0165	-28.1394